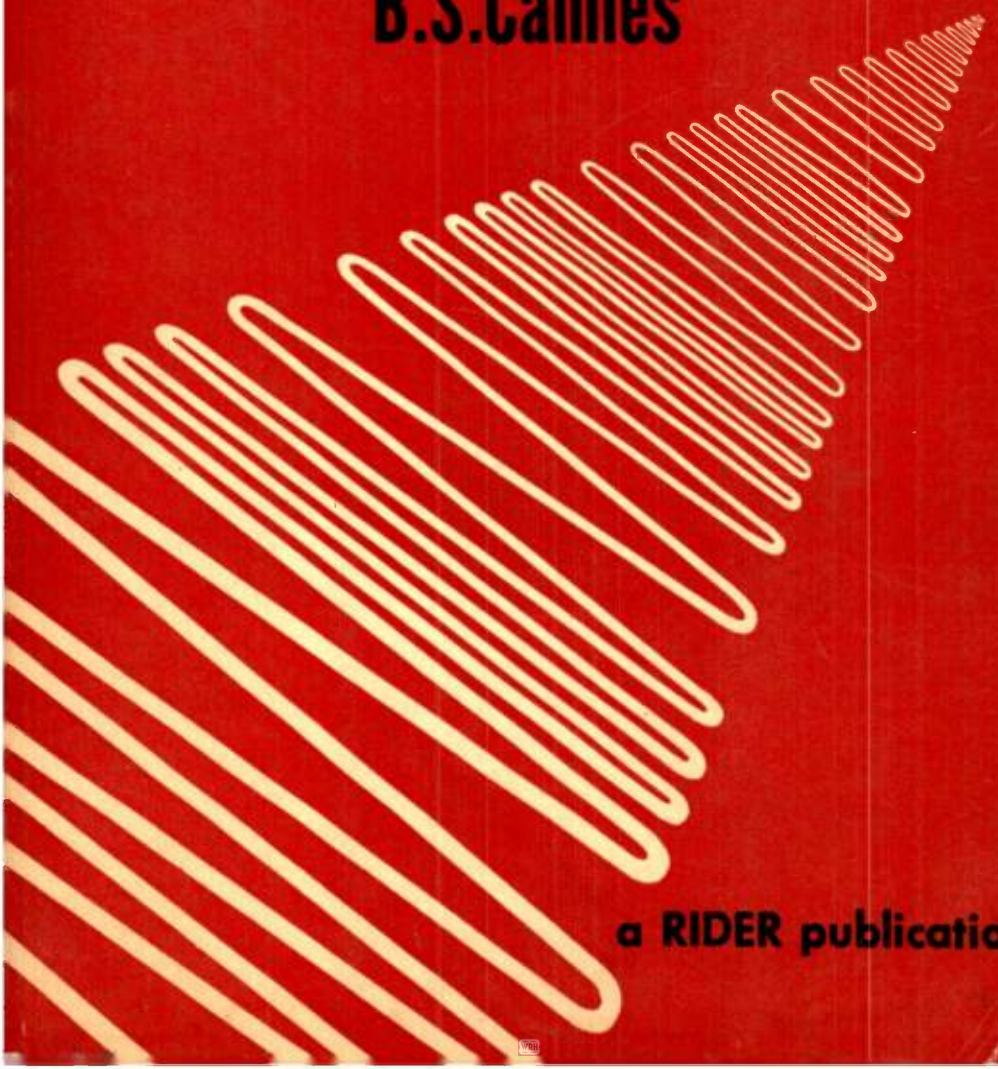
No. 223

PRINCIPLES OF FREQUENCY MODULATION

B.S.Camies



a RIDER publication

PRINCIPLES OF FREQUENCY MODULATION

Applications in Radio Transmitters and Receivers, and Radar

B. S. Camies

LONDON · ILIFFE & SONS LTD

First Published 1959

Published for "Wireless World" by Iliffe & Sons Limited,
Dorset House, Stamford Street, London, S.E.1. Made and printed
in England at The Chapel River Press, Andover, Hants. BKS 3387

CONTENTS

PREFACE

THIS book is intended primarily for students, radio engineers and radio amateurs. In it I have attempted to give a reasonably comprehensive account of the fundamentals of frequency modulation and its applications.

The theory is given in some detail and a number of calculations are included to illustrate the sideband structure and bandwidth of frequency-modulated waves. Some space is devoted to the relative advantages of f.m. and a.m. receivers in receiving signals in the presence of various kinds of interference.

The second half of the book is devoted to the applications of frequency modulation. Possibly of greatest interest to the intended reader is the design of broadcast f.m. receivers and this subject is given as complete a treatment as possible in a book of this size. Nevertheless, frequency modulators of a number of types and their use in f.m. transmitters are also described, as is also the use of f.m. in microwave links, in radar, in telegraphy and in facsimile transmission.

Numerical calculations are included throughout the book to show how simple design calculations may be performed and to illustrate the practical magnitudes of quantities.

B. S. CAMIES.
15th December, 1958.

ACKNOWLEDGEMENTS

Thanks are due to E. K. Cole Ltd. for permission to publish the circuit diagram of the a.m.-f.m. receiver featured in Fig. 6.10 and to H. J. Leak and Co. Ltd. for permission to publish the circuit diagram of the f.m. tuner featured in Fig. 6.12.

1

BASIC PRINCIPLES OF FREQUENCY MODULATION

Introduction

THE first step in effecting communication by means of radio between one point and another is to radiate a carrier wave from a transmitter at one point and to intercept it at the other. This establishes a link between the two points: to send messages over the link the carrier wave must be modified in some way, i.e. must be modulated.

Two properties of a carrier wave which can be varied to convey messages are the amplitude and the frequency of the wave. In the early days of radio communication, messages were sent using the principles of telegraphy, the carrier wave being interrupted, or its frequency displaced from one value to another, to form the dots and dashes of the Morse code. These are early examples of amplitude and frequency modulation in which the modulating signals are rectangular waves.

LIMITATIONS OF AMPLITUDE-MODULATED BROADCASTING

When broadcasting began in the early 1920's the technique of frequency-modulating a carrier wave by an audio signal had not been developed and programmes were radiated by amplitude modulation of medium and long waves: even today most broadcasting is carried out by this means. Amplitude modulation has, however, serious limitations. Perhaps the most serious of these is that a receiver designed for amplitude-modulated signals also responds to a large variety of unwanted signals known collectively as interference. These signals can be classified into two main categories:

(1) man-made interference, comprising signals radiated by electrical equipment such as electric motors, neon signs and switches. Man-made interference is particularly troublesome

in large built-up areas such as cities because of the large amount of electrical equipment present and because of the difficulty of providing receivers with aerials which are clear of the interfering fields.

(2) natural interference, e.g. signals generated by lightning flashes.

A second disadvantage of medium- and long-wave broadcasting is the limited fidelity of reproduction possible. To accommodate all the transmissions required in the wavebands available, the carrier spacing of broadcasting stations is fixed by international agreement at 8, 9 or 10 kc/s, the precise figure depending on the waveband and the geographical location of the stations. To avoid interference between such closely-spaced signals, receivers must be selective and their audio-frequency response falls off steeply above approximately 3 kc/s. Fortunately the wavelength allocation is usually such that stations with adjacent frequencies are remote geographically and a receiver with a suitably-wide passband can give a more extended frequency response from a strong local station. This is possible because signals on adjacent carrier frequencies originate from transmitters at a great distance and are usually too weak to cause serious interference.

Such better quality of reproduction is, however, only possible during daylight hours when the range of medium-wave signals is limited to, say, 200 miles. After dark these signals are reflected at ionised layers in the upper atmosphere and can travel up to 1,500 miles or more, causing interference to high-quality reception as far away as this. Even if the interfering signal is weak enough for its modulation to be inaudible, there is a beat note of 8, 9 or 10 kc/s between the wanted and interfering carriers which is particularly annoying. Except in specialised receivers, therefore, the passband is made narrow to minimise adjacent-channel interference, thus limiting the standard of reproduction as mentioned earlier.

ADOPTION OF FREQUENCY MODULATION IN AMERICA

In an effort to improve reception and quality of reproduction of local programmes for city dwellers, the U.S.A. introduced frequency-modulated stations in 1939, this being the first time that frequency modulation was employed for broadcasting. Credit for the introduction of this system is primarily due to Major E. H. Armstrong who had been interested in frequency modulation since 1914.

By 1936 he had developed his system to a state where it could be successfully employed in broadcasting and in that year he published a paper demonstrating the reduction of interference that was possible using this system.

Three years later the first f.m. transmitters began operation, and by 1951 there were more than 600 f.m. transmitters operating in the 88–108 Mc/s band in America.

ADOPTION OF FREQUENCY MODULATION IN GREAT BRITAIN

Medium-wave broadcasting in Europe deteriorated markedly after the Second World War as a result of the increase in man-made appliances generating interference and more particularly because of the steady growth in the number of medium-wave stations. In Great Britain it was decided that a v.h.f. service operating in the 87·5–100 Mc/s band was the only way to provide a first-class distribution of home programmes, and after a long series of tests using frequency modulation and amplitude modulation a decision was reached in favour of frequency modulation. In May 1955 the first B.B.C. f.m. station began operation at Wrotham, Kent, giving good reception of the Home, Light, and Third programmes over the South-East corner of Britain. Since then further f.m. transmitters have been completed and the f.m. service now covers the whole country.

FUNDAMENTAL FEATURES OF A V.H.F. SERVICE USING FREQUENCY MODULATION

It was originally thought that, by use of a small frequency swing, an f.m. transmission could be made to occupy less bandwidth than an amplitude-modulated transmission, but this was disproved by Carson in 1922 who showed that the bandwidth for f.m. was at least twice the highest modulating frequency. It follows, therefore, that the bandwidth for f.m. is, in general, greater than for amplitude modulation for a given modulating frequency. In fact, as shown by Armstrong in 1936, if full advantage is to be taken of the noise reduction possible with frequency modulation, the bandwidth occupied by a transmission modulated up to 15 kc/s is of the order of 240 kc/s. Such a bandwidth is tolerable only on a v.h.f. range and the British f.m. transmitters operate in Band II (87·5–100 Mc/s).

Waves with frequencies of this order have a number of properties (listed below) which have an important bearing on the performance

of a v.h.f. broadcasting service. These are some of the factors which had to be considered in making the choice between f.m. and a.m. after the decision was taken to establish a v.h.f. broadcasting service in this country.

(1) V.h.f. waves have a range limited to approximately 100 miles, except during conditions of anomalous propagation. Such a range is suitable for distribution of domestic programmes giving little likelihood of interference from v.h.f. transmitters in neighbouring countries. The effects of such little interference which may occur is negligible if frequency modulation is used in preference to amplitude modulation. In f.m. reception strong signals tend to swamp and obliterate weaker ones completely, this being known as the *capture effect* which is described in Chapter 3.

Experiments show that an f.m. transmitter has a range of four or five times that of an a.m. transmitter with the same power output, range being determined for equal signal-noise ratio.

(2) Motor-car ignition systems radiate appreciable energy in the v.h.f. band (30–300 Mc/s) and this can cause interference to receivers operating in these bands. If no means of combating this existed, there would be very little point in changing from medium-wave to v.h.f. transmission because this would only exchange one form of interference for another. It is difficult to eliminate ignition interference from a.m. receivers (e.g. television receivers) but a well-designed f.m. receiver can be made virtually immune to amplitude modulation of received signals and this practically eliminates ignition and many other forms of interference.

(3) The wavelength range of the v.h.f. band is from 1–10 metres. The frequencies used for f.m. transmission are around 90 Mc/s for which the wavelength is just over 3 metres. This is so small that it is practicable to construct compact highly-directive aerials for transmission and reception and these can easily be supported on masts or chimney tops. Such aerials offer a number of important advantages.

A directive transmitting aerial considerably reduces the transmitter power output necessary to produce a given service area. For example, an aerial with 6 dB gain requires an output of 25 kW to produce the same field as a 100 kW

transmitter feeding a simple dipole aerial. This economy arises because the directive aerial reduces radiation at near vertical directions (which would serve no useful purpose) and concentrates it in the horizontal plane where the energy is wanted.

A directive receiving aerial is also advantageous because it effectively increases the magnitude of the received signal. However, a directive aerial is frequently used not because of its ability to increase the magnitude of wanted signals but because it can markedly reduce the magnitude of unwanted signals. In general, directive aerials have a minimum response to signals arriving in a particular direction, and if this direction does not coincide with that of the wanted signal also, it is possible to orient the aerial so as to give minimum response to the unwanted signals. This may not be the orientation which gives maximum response to the wanted signals but it usually gives best results. The interfering signals may be from car ignition systems or from the transmitter of the wanted signals, signals reaching the receiving aerial via a reflection at some large object such as a gasholder or a hill. The effect of reflections is discussed more fully in the next section (4).

(4) Perhaps the most serious disadvantage of v.h.f. waves is that they are readily reflected from large objects such as buildings, gasholders, hills, trees or even aircraft. Thus v.h.f. waves can be received at an aerial via any number of paths, the shortest being along the straight line connecting the transmitting to the receiving aerial. The other paths are longer because they include one or more reflections at objects such as those mentioned. Because of the difference in path length, the direct and indirect waves are not, in general, in phase, and if one of the reflecting objects is an aircraft and can move, there can be variations in phase difference from moment to moment which can cause violent fluctuations in received signal strength. The variations in image brightness on a television receiver caused by an aircraft flying nearby are well known and are commonly termed *aircraft flutter*.

In an a.m. sound receiver such variations in signal strength can cause annoying variations in volume: in an f.m. receiver the variations in amplitude caused by multipath reception are equivalent to amplitude modulation of the received

carrier and can virtually be eliminated by effective limiting in the receiver. However, in multi-path reception the amplitude modulation is accompanied by phase modulation which can give rise to distortion of the a.f. output. In a poorly-designed receiver this distortion can be severe, but can often be reduced by using a better aerial, preferably a directive one. In a well-designed receiver with effective limiting, multi-path distortion can usually be reduced to very small and often negligible proportions. Although multi-path distortion is most troublesome in the fringes of the service area of an f.m. transmitter it is also found in regions of high field strength.

F.M. TRANSMITTERS

If a broadcast transmitter is to operate in the v.h.f. band the aerial system can be made directive without being unwieldy in size. This is possible even if the transmitter is required to be omni-directional, for the aerial can be designed to minimise upward and downward radiation, concentrating energy flow in the horizontal plane containing the aerial. By this means it is possible for a given power input into the aerial to obtain, say, a four-fold increase in the signal strength at a given distant point. Alternatively, to give a wanted field strength at a given point, the use of a directive aerial reduces the required transmitter power output. Thus a 50 kW transmitter with a directive aerial can give the same service area as a 200 kW transmitter with a simple dipole aerial. Thus a v.h.f. broadcast transmitter does not need to give as much power output as, say, a medium-wave transmitter, and this economy is obtained whether the v.h.f. transmitter is amplitude-modulated or frequency-modulated. It is purely a function of the carrier frequency.

The modulation of an a.m. transmitter may be impressed on the carrier wave in the final stage or in an earlier stage. If modulation is achieved in the final stage (termed *high-power modulation*) all the r.f. amplifiers including the final (modulated) one can operate in class C. Such operation is in effect very efficient, converting 80 per cent of the power taken from the h.t. supply into useful power output.

However the modulator must supply considerable a.f. power (50 kW for a 100 kW transmitter) and to design an amplifier to supply such a large power output with very low distortion is not easy. The amplifier is inevitably expensive, bulky and of limited efficiency.

An alternative system is to modulate the carrier wave at an early stage and to amplify the modulated r.f. signal to the required power afterwards. This is known as *low-power modulation* and it avoids the necessity for a large a.f. amplifier. However, the design of modulated r.f. amplifiers for a low-power modulated transmitter is difficult because they must be linear to avoid distortion of the modulation envelope. The efficiency of such amplifiers cannot be high for they must operate in class B. There is thus little to choose between high-power and low-power modulation systems and both are in common use in a.m. transmitters.

These considerations do not apply to f.m. transmitters in which modulation is achieved at an early stage because little a.f. power is necessary and small receiving-type valves can be used. The frequency-modulated signals from the modulator are of constant amplitude and can be amplified to any desired power level by highly-efficient class-C stages. Thus all the r.f. stages in an f.m. transmitter can operate in class C and no large a.f. amplifier is necessary. As a result an f.m. transmitter can be more efficient and therefore much smaller than an a.m. transmitter of the same power output.

F.M. RECEIVERS

Although equipment for transmitting f.m. signals is more efficient and smaller than that for a.m. signals, receiving equipment is necessarily more complex and more expensive. The reasons for this are given in detail in Chapters 5 and 6 but can be briefly summarised as follows. Due to reflection and absorption in buildings and ground, v.h.f. signal strengths can vary over a surprisingly large range within distances of a few feet. At an unfavourable receiving site it is possible for the signal even from a nearby v.h.f. transmitter to be less than 10 μV/metre. Moreover, if more than one propagation path exists, the amplitude of this signal can rise and fall. To give good reproduction of such a signal, an f.m. receiver must be capable of efficient limiting even when the input signal is well under 10 μV. Limiting is usually achieved by use of a device with a non-linear characteristic and the circuits commonly employed require input signals of several volts for satisfactory performance.

Thus an f.m. receiver requires considerable gain (at least 10^6) before the limiter. It would be difficult to obtain such gain at the carrier frequency. Moreover, it is scarcely practicable to design f.m. detectors to operate at such frequencies. F.m. receivers are therefore superheterodynes with an intermediate frequency around

10·7 Mc/s, at which conventional valves can give good gain and satisfactory discriminators can be built. Even so, the large pre-limiter gain necessitates more stages in an f.m. receiver than in a medium-wave receiver. The adoption of the superheterodyne principle also introduces the problem of ensuring oscillator stability. Wandering of the oscillator frequency causes mistuning of the receiver, and a high degree of stability is necessary to reduce such mistuning to acceptable proportions. Some manufacturers include a reactance-valve to provide automatic frequency correction: this introduces yet a further stage into the receiver. Thus f.m. receivers are inevitably more complex than a.m. receivers.

We have so far discussed f.m. as a means of high-quality sound broadcasting, but it is used for other purposes also. The sound accompaniment in American television services employs f.m., and the microwave links used by television authorities to send vision signals from point to point, e.g. from a television outside-broadcast site to a control room, also employ f.m. Certain radar equipments and aircraft altimeters also make use of f.m. signals.

It is the purpose of this book to establish the reasons for the superiority of f.m. over a.m. and to give the principles of operation of the equipment used to generate and to receive f.m. signals.

2

THEORY OF FREQUENCY MODULATION

Introduction

IN this chapter we shall develop an expression for a frequency-modulated wave from which we can deduce the side-band structure and the passband necessary for transmission.

First, however, we shall consider the modulation of a carrier wave in general. Modulation is necessary to enable a carrier wave to convey a message from one point to another. It is achieved by varying a property of the carrier wave in accordance with the waveform of the signal to be transmitted. The latter is usually known as the modulating signal and in practice may be a Morse signal as in telegraphy, an audio signal as in sound broadcasting, a video signal as in television broadcasting, or a facsimile signal as in picture transmission.

A carrier wave has three properties which can be varied to effect modulation, namely its amplitude, its frequency, or its phase. Thus we can distinguish three types of modulation: amplitude modulation (a.m.), frequency modulation (f.m.) and phase modulation (p.m.). Although all three types of modulation can occur simultaneously, practical modulators are usually designed to achieve one type and to minimise the other two.

The equation to a carrier wave can be written

$$e = A \cos (\omega t + \phi)$$

where e = instantaneous value of the carrier at a time t

A = amplitude of the carrier

ω = angular frequency of the carrier $= 2\pi f$ where f is the frequency

ϕ = phase angle of the carrier, i.e. the angle when $t = 0$.

The problem in effecting modulation is to vary A, f or ϕ in accordance with the waveform of the modulating signal, so that variations in the amplitude and the frequency of the modulating

signal are faithfully impressed upon the carrier wave, impressed moreover in such a way that the modulating-signal waveform can be extracted from the modulated carrier wave by relatively simple means in the detector stage of a receiver.

DISTINCTION BETWEEN AMPLITUDE, PHASE AND FREQUENCY MODULATION

In amplitude modulation the carrier amplitude (A) is swung above and below its average or unmodulated value at a frequency equal to that of the modulating signal. The extent of the upward and downward changes in carrier amplitude is proportional to the amplitude of the modulating signal. This is illustrated in Fig. 2.1 which illustrates the waveform of the modulated carrier for small-amplitude and large-amplitude modulating signals, represented for simplicity as sinusoidal in form.

From Fig. 2.1 we can see that if the amplitude of the modulating signal is continually increased, a point will be reached at which the carrier amplitude swings between twice its average unmodulated amplitude and zero. Any attempt to increase the amplitude of the modulating signal beyond this point results in periods of zero carrier amplitude. If such a carrier wave is applied to a detector the modulating-frequency output is inevitably distorted. This limit to the degree of modulation which can be impressed upon the carrier is one of the disadvantages of amplitude modulation.

In phase modulation the phase angle (ϕ) is swung above and below its average or unmodulated value at the frequency of the modulating signal. The extent of the phase swing is proportional to the amplitude of the modulating signal. The resulting phase-modulated carrier is similar to a frequency-modulated carrier and, in fact, transmitters can be designed to produce f.m. or p.m. by inclusion of suitably-designed frequency-discriminating networks in the modulator chain. The relationship between f.m. and p.m. will be made clear in the following text.

In frequency modulation the frequency (f) of the carrier wave is swung above and below its average or unmodulated value at the frequency of the modulating signal. This is illustrated in Fig. 2.2 for small-amplitude and large-amplitude modulating signals. Notice that the amplitude of the frequency-modulated wave is the same as for the unmodulated wave.

Normally the changes in frequency due to modulation are very small (e.g. less than 1/1,000th of the carrier frequency) and there is

no theoretical limit to the maximum depth of modulation as in amplitude modulation. Increase in the amplitude of the modulating signal can always cause greater swings of the carrier frequency. This may in practice cause distortion because the increased frequency swings may generate sidebands lying outside the passband

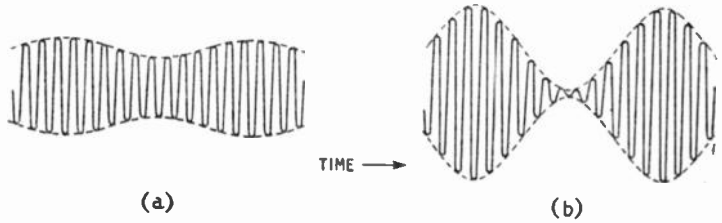


Fig. 2.1. Amplitude modulation for (a) a small-amplitude, and (b) a large-amplitude modulating signal of sinusoidal waveform

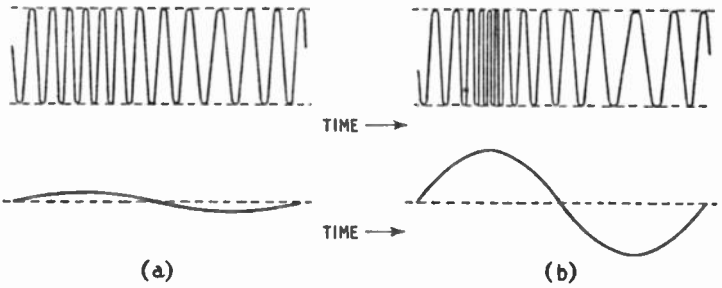


Fig. 2.2. Frequency modulation for (a) a small amplitude and (b) a large-amplitude modulating signal of sinusoidal waveform

of the transmitter or receiver circuits. If any significant sidebands are lost or greatly attenuated in this way, harmonic distortion of the detected signal will occur.

AMPLITUDE MODULATION

It might be thought that amplitude modulation could be achieved by making the amplitude of the carrier wave proportional to the amplitude of the modulating signal. In such a system A is replaced by $mA \cos pt$ where p is the angular frequency of the modulating signal and m is its amplitude. Amplitude modulation of this type may be possible, but it would require a detector of a complex type

to eliminate distortion. Detection is simpler if the *deviation* of the carrier wave from its unmodulated value is made proportional to the modulating signal. This is illustrated in Fig. 2.3, in which shallow sinusoidal modulation is represented.

A is replaced by $A(1 + m \cos pt)$ in amplitude modulation of this type, and the full expression for the amplitude-modulated wave becomes

$$e = A(1 + m \cos pt) \cos \omega t$$

in which the carrier phase angle (ϕ) is omitted because it has no significance in amplitude modulation. m expresses the amplitude of the modulating signal and its maximum value is unity. This can easily be demonstrated by putting $pt = 0$, which gives $m \cos pt = 1$, and the carrier amplitude is 2A. When $pt = \pi$ rads, $m \cos pt = -1$, and the carrier amplitude is zero. A value of m of 1 thus causes the carrier amplitude to swing between 0 and 2A, and this is the deepest modulation possible without distortion. m is commonly termed the *modulation factor* and is often given in the form of a percentage, e.g. when $m = 0{\cdot}5$ this is expressed as 50 per cent modulation.

The expression for the modulated carrier can be expanded thus

$$e = A \cos \omega t + Am \cos \omega t \cos pt$$

$$= A \cos \omega t + \frac{mA}{2} \cos (\omega - p)t + \frac{mA}{2} \cos (\omega + p)t$$

The first of these terms is the original carrier which is present at constant amplitude in spite of amplitude modulation. The second term represents a sinusoidal quantity with an angular frequency

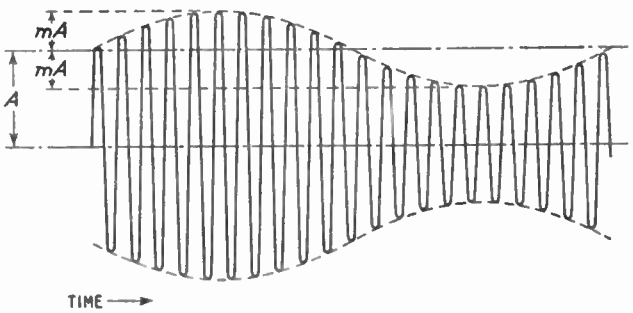


Fig. 2.3. Illustrating the definition of modulation index in amplitude modulation

$(\omega - p)$ equal to the difference between carrier and modulating angular frequencies and with an amplitude $mA/2$ proportional to the modulation index. The third term represents a sinusoidal quantity with an angular frequency $(\omega + p)$ equal to the sum of the carrier and modulating angular frequencies and with an amplitude $mA/2$

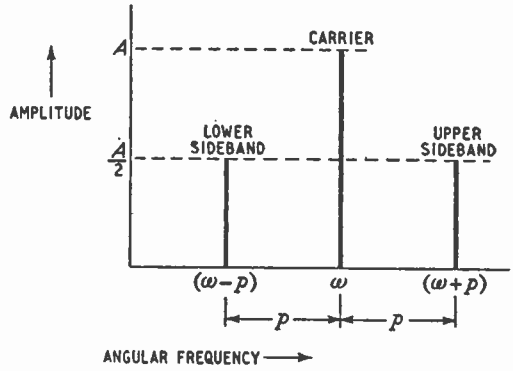


Fig. 2.4. Spectrum of 100 per cent amplitude-modulated wave

also proportional to the modulation factor. The second and third terms are due to modulation (for they vanish when m is put equal to 0) and are known as *sidebands*. For 100 per cent modulation the amplitude of the sidebands is one half that of the carrier, and the spectrum of the wave can be represented as in Fig. 2.4.

When the modulating signal is complex, each sinusoidal component gives rise to a pair of sidebands symmetrically disposed about the carrier frequency and displaced from it by the modulating frequency. The frequency range occupied by the sidebands is equal to twice the highest modulating frequency, i.e. is 6 kc/s for a telephony transmitter for which the modulating frequencies are limited to 3 kc/s, and to 30 kc/s for a high-quality broadcast transmitter for which the modulating frequencies extend to 15 kc/s.

PHASE MODULATION

The expression for a carrier wave phase-modulated by a sinusoidal modulating signal may be written

$$e = A \cos (\omega t + m\phi_d \cos pt)$$

in which the quantity $m\phi_d \cos pt$ replaces ϕ in the expression for an unmodulated wave. The phase angle varies sinusoidally at an

angular frequency of p. m is a factor expressing the amplitude of the modulating signal and may have any value between 0 and 1. For modulating signals of maximum amplitude $m = 1$ and the values of the phase angle ϕ for various values of pt are as follows

Values of pt (rads)	*Values of* cos *pt*	*Values of* $m\phi_d \cos pt$ *(rads)*
0	1	ϕ_d
$\pi/2$	0	0
π	-1	$-\phi_d$
$3\pi/2$	0	0
2π	1	ϕ_d

This shows that the phase angle swings between the limits of ϕ_d and $-\phi_d$. ϕ_d is therefore the *peak phase deviation.* It is significant that the extent of the phase deviation for a given value of m does not depend on the frequency of the modulating signal, i.e. a particular a.f. input voltage to a p.m. transmitter produces the same peak values of phase shift at 50 c/s as at 10,000 c/s. This is not true for a frequency-modulated transmitter.

FREQUENCY MODULATION

In frequency modulation the carrier frequency is swung about its mean value $f_0 (= \omega/2\pi)$ between the limits $(f_0 + f_d)$ and $(f_0 - f_d)$. The extent of the swing is proportional to the modulating-signal amplitude, and we will assume that f_d represents the peak swing reached on maximum values of modulating-signal amplitude: f_d is known as the *deviation* and is commonly 75 kc/s in f.m. broadcasting services. If the modulating signal has a sinusoidal waveform of angular frequency p, the instantaneous carrier frequency f is given by

$$f = f_0 + f_d \cos pt$$

If the modulating-signal amplitude is not at its maximum value, it can be written as m times the maximum, and the instantaneous frequency is then given by

$$f = f_0 + mf_d \cos pt$$

where m may have any value between 0 and 1. m is not known as the modulation factor in frequency modulation: it does, however,

enter into the expression for the modulation factor which is defined later.

The equation to the carrier wave is

$$e = A \cos \theta$$

When the frequency is constant, $\theta = \omega t$ and the equation becomes

$$e = A \cos \omega t$$

When, as in frequency modulation, the carrier frequency varies, the value of θ must be obtained by integration thus

$$\begin{aligned}
\theta &= \int \omega . dt \\
&= 2\pi \int f . dt \\
&= 2\pi \int (f_0 + mf_d \cos pt)\, dt \\
&= 2\pi \left(f_0 t + \frac{mf_d}{p} \sin pt\right) \\
&= \omega t + \frac{m2\pi f_d}{p} \sin pt \\
&= \omega t + \frac{mf_d}{f_m} \sin pt
\end{aligned}$$

where f_m is the modulating frequency and is equal to $p/2\pi$.

Thus the equation to the frequency modulated wave is

$$e = A \cos \left(\omega t + \frac{mf_d}{f_m} \sin pt\right)$$

This expression is of the same form as that for a phase-modulated wave, but the peak phase swings for the frequency-modulated wave are not independent of frequency as in phase modulation. The phase swings are given by mf_d/f_m and are proportional to the modulating-signal amplitude (as in phase modulation), and inversely proportional to the modulating frequency. For a given a.f. modulating voltage into an f.m. transmitter, therefore, the phase swing for a 50 c/s signal is 200 times that for a 10,000 c/s signal.

If, therefore, a modulating signal is applied to a network whose output (for constant input) in inversely proportional to frequency, and if this output is then passed to a phase modulator, the result is a

frequency-modulated wave. The required network has an attenuation falling at 6 dB per octave throughout the audio range. Alternatively, if a modulating signal is applied to a network with an output directly proportional to frequency, and if this output is then passed to a frequency modulator, the output is a phase-modulated wave.

SIDEBAND STRUCTURE OF A FREQUENCY-MODULATED WAVE

The equation to a frequency-modulated wave may be written

$$e = A \cos (\omega t + M \sin pt)$$

where M is the *modulation index* and is given by

$$M = \frac{mf_d}{f_m}$$

The modulation index is directly proportional to the amplitude of the modulating signal and to the deviation, but is inversely proportional to the modulating-signal frequency.

Expanding the above expression we have

$$e = A \cos \omega t \cos (M \sin pt) - A \sin \omega t \sin (M \sin pt).$$

This may be written in the form

$$\begin{aligned} e/A = {} & \mathcal{J}_0(M) \cos \omega t + \mathcal{J}_1(M)[\cos (\omega + p)t - \cos (\omega - p)t] \\ & + \mathcal{J}_2(M)[\cos (\omega + 2p)t - \cos (\omega - 2p)t] + \ldots \\ & + \mathcal{J}_{2n-1}(M)[\cos (\omega + \overline{2n-1}p)t - \cos (\omega - \overline{2n-1}p)t] \\ & + \mathcal{J}_{2n}(M)[\cos (\omega + 2np)t - \cos (\omega - 2np)t] \end{aligned}$$

in which $\mathcal{J}_0(M)$, $\mathcal{J}_1(M)$, etc. are Bessel functions of M of the first kind. The values of these functions can be obtained from tables. For example, the value of $\mathcal{J}_2(5)$ is 0·0466.

Examination of this rather cumbersome expression for a frequency-modulated wave shows that it contains a number of components with angular frequencies equal to ω, $(\omega + p)$, $(\omega - p)$, $(\omega + 2p)$, $(\omega - 2p)$, etc. The first of these components is at carrier frequency, the second and third constitute a sideband pair symmetrically disposed about the carrier frequency and displaced from it by a frequency equal to the modulation frequency. The fourth

and fifth constitute a pair of sidebands, also symmetrically disposed about the carrier frequency, but displaced from it by twice the modulation frequency. There are a large number of such pairs of sidebands displaced from the carrier frequency by multiples of the modulation frequency. Thus, a single sinusoidal modulating signal can give rise to a large number of pairs of sidebands: in amplitude modulation such a modulating signal produces only a single pair of sidebands.

The two sidebands constituting each pair in a frequency-modulated wave have the same amplitude, but this is not linearly related to the modulating-signal amplitude as in amplitude modulation but is related to it according to a Bessel function of M. The amplitude of the carrier-frequency component is also dependent

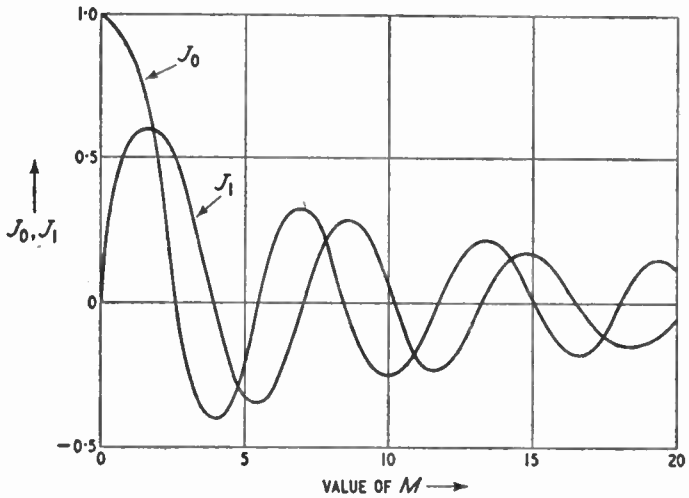


Fig. 2.5. Variation of carrier amplitude (J_0) and first sideband pair amplitude (J_1) with modulation index

on a Bessel function (J_0) of M. The way in which the amplitude of the carrier and first pair of sidebands depends on M is illustrated in Fig. 2.5.

Consider first the value of the Bessel function J_0 which determines the carrier amplitude. This has a value of unity when the modulation index is zero: that is to say the carrier has a certain amplitude in the absence of modulation. As M is increased the curve of $J_0(M)$ decreases and then swings through the horizontal

axis at a steadily decreasing amplitude with a shape similar to that of a damped wavetrain. This shows that the carrier amplitude varies with the modulation index and, in fact, is zero for a number of values of M: this contrasts strongly with amplitude modulation in which the carrier amplitude is constant and independent of modulation.

The curve of J_1 has a similar shape, but starts from zero when M is zero, as indeed it must, for the amplitude of the first pair of sidebands must clearly be zero when there is no modulation. Again, there are values of M for which the amplitude of these sidebands is zero.

The curves for J_2, J_3, etc. all start at zero when $M = 0$, and have the same general shape as the curve of J_1.

Since $M = mf_a/f_m$, if the frequency of the modulating signal is fixed, M is proportional to the modulating-signal amplitude. Thus, the curves of J_0 (or J_1), etc. illustrate the variation in the amplitude of the carrier (or first sideband pair) as the modulating-signal amplitude is increased.

If the amplitude of the modulating signal is fixed, M is inversely proportional to the modulating-signal frequency. Thus the curves of J_0, J_1, etc. also represent the variation in the amplitude of the carrier component (or first sideband pair) as the modulating-signal frequency is decreased.

Practical values of the modulation index vary considerably with frequency. For example, consider a modulating frequency of 15 kc/s and let the deviation be 75 kc/s.

We thus have

$$M = \frac{mf_a}{f_m}$$

$$= \frac{75m}{15}$$

$$= 5m$$

The upper limit of m is 1 and M cannot exceed 5. If we determine the values of the Bessel functions J_0, J_1, etc. for a value of M of 5, we can calculate the amplitude of the carrier, first sideband pair, second sideband pair, etc. These are plotted in Fig. 2.6, which shows that the amplitude of the sideband pairs decreases as the order increases and, in fact, beyond the eighth sideband pair are less than 1 per cent of the unmodulated carrier amplitude and small

enough to be neglected. Thus at 15 kc/s there are eight significant pairs of sidebands at maximum modulating-signal amplitude and correspondingly fewer for smaller signal amplitudes.

Now consider a modulating-signal frequency of 50 c/s, the deviation being still 75 kc/s.

$$M = \frac{m.75 \times 10^3}{50}$$

$$= 1{,}500m$$

If m is 1, M is 1,500. If we determine the Bessel functions J_0, J_1, etc., for a value of M of 1,500 we find that there are hundreds of

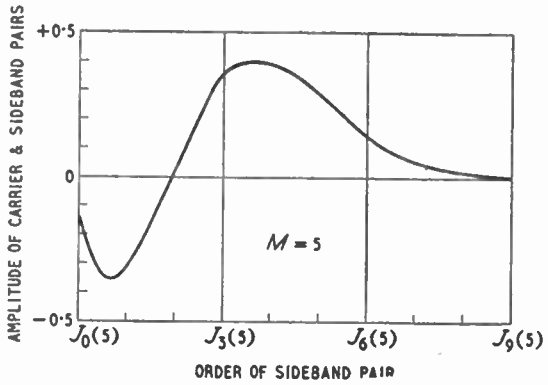


Fig. 2.6. Curve illustrating relative amplitudes of carrier and sideband pairs for a modulation index of 5

pairs of sidebands of significant amplitude. As in Fig. 2.6 their amplitude decreases with increase of the order. The number becomes smaller if the modulating-signal amplitude is reduced, but m must be 1/300 ($\frac{1}{3}$ per cent of maximum amplitude) to make M 5, giving eight pairs of sidebands as shown above.

These calculations show that the number of pairs of significant sidebands is roughly in direct proportion to the value of M and therefore increases with increase in modulation-signal amplitude (if frequency is constant) and with decrease in modulation-signal frequency (amplitude remaining constant). If the amplitude and frequency of a modulating signal are increased in the same ratio, the value of M remains unchanged and the number of sideband pairs remains the same. The relative amplitude of the carrier and

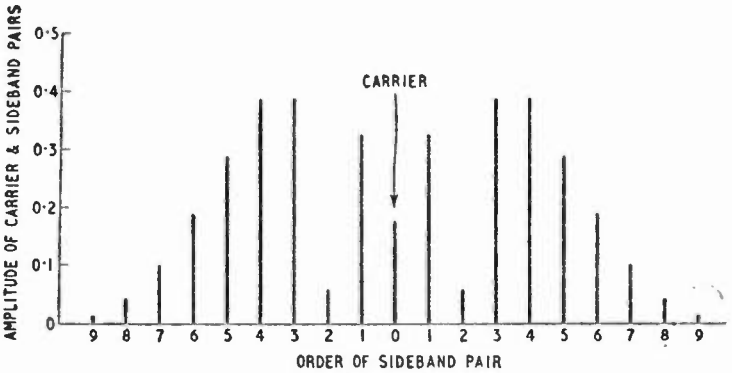


Fig. 2.7. Spectrum of frequency-modulated wave for a modulation index of 5

sidebands is the same, giving the same spectrum pattern, but the sideband spacing is greater because of the increased modulation frequency.

A typical spectrum pattern for a frequency-modulated wave is given in Fig. 2.7: several pairs of sidebands have an amplitude greater than that of the carrier.

BANDWIDTH OCCUPIED BY A FREQUENCY-MODULATED WAVE

It was shown above that there were eight significant pairs of sidebands for a modulating signal of maximum amplitude at 15 kc/s. The edges of the bandwidth occupied by such a signal are marked by the outermost (eighth) pair of sidebands and the bandwidth is given, therefore, by

$$\text{bandwidth} = 2 \times 8 \times 15 \text{ kc/s}$$
$$= 240 \text{ kc/s}.$$

For a 50 c/s signal the number of significant sidebands is much greater, but because they are spaced at 50 c/s intervals (compared with the 15 kc/s intervals above) the frequency range occupied by the sidebands is approximately the same. One of the most useful features of frequency modulation is that constant-amplitude modulating signals give a constant bandwidth irrespective of the modulation frequency. In phase modulation the bandwidth increases linearly with the modulating frequency.

The bandwidth can be determined, as in the above calculation, by considering modulation by a peak amplitude signal at the highest modulating frequency.

Since $m = 1$

$$M = \frac{f_d}{f_m}$$

and is equal to the ratio of the deviation to the highest modulating frequency.

This is known as the *deviation ratio* and it, together with the value of the deviation, determines the passband required. The relationship between deviation ratio and passband is illustrated in

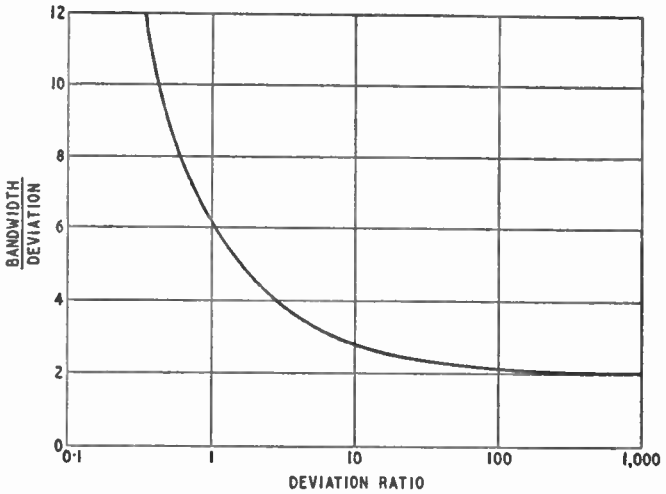


Fig. 2.8. Relationship between bandwidth, carrier frequency swing and modulation index for a frequency-modulated wave

Fig. 2.8 which was evaluated from tables of Bessel functions. The passband is expressed in terms of the deviation. The curve shows that if the deviation ratio is equal to 5 the ratio of passband to deviation is 3·4.

The passband is thus given by

$$3{\cdot}4 \times 75 \text{ kc/s} = 240 \text{ kc/s}$$

which confirms the result deduced above.

Communications f.m. systems such as those used by the police and other organisations commonly employ a maximum modulating frequency of 3 kc/s with a deviation of ± 15 kc/s. The deviation ratio is again 5 and the passband is given by

$$3{\cdot}4 \times 15 \text{ kc/s} = 51 \text{ kc/s}$$

These values of passband are significantly greater than the peak frequency swing. For high-quality broadcasting the passband is 240 kc/s compared with a peak frequency swing of 150 kc/s and for communications systems the passband is 51 kc/s compared with a peak frequency swing of 30 kc/s.

3

FREQUENCY MODULATION AND INTERFERENCE

Introduction

AS pointed out in Chapter 1, the chief problem in radio reception is that of reducing or eliminating interference, and in this respect frequency modulation is, in general, superior to amplitude modulation. In this chapter we shall compare in detail the behaviour of an f.m. and an a.m. receiver towards the types of interference commonly encountered. This will account for the good performance of the f.m. receiver, but will also show that the full advantages of frequency modulation can be obtained only by use of well-designed receivers. The design of an f.m. receiver is, in fact, more complex than that of an a.m. receiver and this, of course, makes it necessarily more expensive.

FORMS OF INTERFERENCE

Co-channel Interference

Interference in its widest meaning embraces all forms of unwanted signal or noise. Possibly the most familiar form of interference is that due to an unwanted signal whose modulation can be heard in addition to that of the wanted signal. If the unwanted signal has the same or substantially the same carrier frequency as that of the wanted signal, the interference is referred to as *co-channel interference*. A special example of co-channel interference occurs in multi-path reception: here the modulation of the unwanted carrier is the same as that of the wanted carrier but is delayed with respect to it, and the delay causes the distortion which so often accompanies multi-path reception.

Adjacent-channel Interference

If there is a substantial difference between the carrier frequencies, the interference is termed *adjacent channel interference*. The effects of

adjacent-channel interference depend on the frequency difference and it is sometimes desirable to distinguish between the interference experienced when the frequency difference is audible (say less than 15 kc/s) and that which occurs when the frequency difference is greater than this. A special case of co-channel or adjacent-channel interference occurs when the unwanted carrier is unmodulated.

Such interference occurs when a receiver picks up, in addition to the wanted signal, an output at the fundamental frequency or at a harmonic of the oscillator of a nearby superheterodyne receiver.

The second main type of interference is that due to noise. This term includes a number of different forms of signal. The click caused by a faulty electric-light switch and the staccato effect produced by a car ignition system are both examples of impulsive noise.

They are characterised by very steep wavefronts and usually the intervals between the wavefronts are long compared with the duration of each pulse.

Impulsive Noise

The spectrum of a noise pulse has a number of discrete components and these are spaced throughout the spectrum at an interval equal to the recurrence frequency of the pulses. In the initial stages of the pulse all the components are in phase and add, so as to give the steep front characteristic of this form of noise. However, not all the components of the pulse are accepted by the passband of the receiver, and the waveform of the pulse signal is modified by the frequency characteristic of the receiver.

Broadly speaking we can say that the output of an i.f. amplifier with a pulse input consists of a damped oscillation at a frequency approximately equal to the mid-band frequency of the amplifier, the duration of the oscillation being approximately equal to twice the period of the bandwidth of the amplifier. For an i.f. amplifier of 200 kc/s bandwidth, a value typical of that for an f.m. receiver, the duration of the oscillation caused by a pulse input is approximately 7 μsec. The oscillation persists longer in a narrow-band than in a wideband amplifier, because of the smaller losses in the highly-selective amplifier.

Moreover, the amplitude of the oscillation is greater in the wideband than in the narrow-band amplifier. A significant feature

of this is that the waveform obtained at the output of an i.f. amplifier bears the characteristics not of the input signal but of the i.f. amplifier itself.

White Noise

A second and quite different type of noise is that which is heard from a sensitive receiver or amplifier when there is no input signal and the gain control is advanced. It is a smooth hiss which is produced by the movements of electrons in conductors, valves or transistors. Unlike impulsive noise, this noise has a continuous spectrum of constant amplitude and by analogy with white light (which also has such a continuous spectrum) is sometimes referred to as *white noise*. It is also known as *fluctuation noise* and *random noise*. White noise is probably the least objectionable form of interference. We shall now list the chief sources of white noise.

Johnson (or Thermal) Noise

Consider an electrical conductor, such as a length of copper wire, which is cut by a plane at right angles to the wire. The free electrons in the wire are in continuous motion at an average velocity which is proportional to the absolute temperature of the conductor. As a result of this motion some electrons cross the plane from left to right whilst others cross it from right to left. An e.m.f. applied to the conductor causes a net flow of electrons through the plane in one direction, this constituting the current due to the e.m.f.

If no e.m.f. is applied to the conductor, the number of electrons crossing the plane from left to right over an appreciable period of time equals the number crossing in the opposite direction. If this were not so, parts of the conductor would either lose or gain electrons causing potential differences to be set up.

However, if we consider the electrons which cross the plane in a very small interval of time it is quite possible for the numbers crossing from left to right to exceed those crossing from right to left: as a result there is a momentary e.m.f. between the two halves of the conductor. Over the next interval of time the balance is likely to be restored by the number of electrons crossing from left to right being less than those crossing the plane in the opposite direction. In other words, the random movement of the electrons sets up a varying e.m.f. between two points on the conductor, the average value of this e.m.f. being zero. The magnitude of this e.m.f. depends on the average electron velocity and thus on the absolute temperature.

The noise voltage V_n is given by

$$V_n^2 = 4\,RKT\Delta f$$

where R is the resistance of the conductor between the points at which the thermal noise voltage is measured,

T is the absolute temperature,

Δf is the bandwidth over which the noise voltage is measured,

and K is Boltzmann's constant $= 1{\cdot}374 \times 10^{-23}$ Joule per °C.

If we substitute the numerical value of Boltzmann's constant and put $T = 290°$ (equal to 17° C and 63° F) we have

$$V_n = 1{\cdot}25 \times 10^{-10}\sqrt{(R.\Delta f)} \text{ volts}$$

showing the noise voltage to be proportional to the square root of the resistance and the bandwidth.

The thermal noise generated in the first stage of a receiver in general causes most interference because it is amplified by all the remaining stages and it is essential to know the order of this voltage. We will therefore calculate the thermal noise voltage likely to be experienced at the grid of the first stage in an f.m. receiver. We will assume the resistance of the tuned input circuit (as damped by the aerial feeder connection) to be 2,000 ohms and the bandwidth of the r.f. circuits to be 200 kc/s. We then have

$$\begin{aligned} V_n &= 1{\cdot}25 \times 10^{-10}\sqrt{(2{,}000 \times 200 \times 10^3)} \text{ volts} \\ &= 1{\cdot}25 \times 10^{-10}\sqrt{(4 \times 10^8)} \text{ volts} \\ &= 1{\cdot}25 \times 10^{-10} \times 2 \times 10^4 \text{ volts} \\ &= 2{\cdot}5 \times 10^{-6} \text{ volts} \\ &= 2{\cdot}5 \text{ microvolts} \end{aligned}$$

Even allowing for the smaller effect which thermal noise has on the performance of an f.m. receiver compared with that which it has on an a.m. receiver it is clear that this sets a lower limit to the signal which can be successfully received.

Shot Noise

The anode current of a valve contains a noise component similar in nature to that of the voltage generated between the ends of a resistance. The anode current is an electron stream composed of a

large number of similar particles. The current appears steady when measured by a meter because the number of electrons arriving at the anode in successive equal intervals of time is equal if the periods are appreciable. However, if the comparison is made over very small intervals of time, the number of electrons per interval does, in fact, vary about the average value. This variation constitutes a noise current and its value I_n is given by

$$I_n^2 = 2I_a e\Delta f$$

where I_a is the anode current

e is the charge on an electron

and Δf is the bandwidth over which the noise current is measured. Substituting the numerical value for the charge e on the electron we have

$$I_n = 5{\cdot}45 \times 10^{-4}\sqrt{(I_a\Delta f)}\ \mu\text{A}$$

which shows that the noise current is proportional to the square root of the anode current and to the square root of the bandwidth.

As a numerical example suppose a valve has an anode current of 5 mA and the bandwidth is 300 kc/s. We have

$$\begin{aligned} I_n &= 5{\cdot}45 \times 10^{-4} \times \sqrt{(5 \times 10^{-3} \times 300 \times 10^3)}\ \mu\text{A} \\ &= 5{\cdot}45 \times 10^{-4} \times 38{\cdot}73\ \mu\text{A} \\ &= 0{\cdot}02\ \mu\text{A} \end{aligned}$$

Partition Noise

In a triode valve substantially all the emission which passes through the control grid arrives at the anode, but in a tetrode valve, a pentode valve or a more complex type of valve the electron stream passing through the control grid is collected by two or more electrodes, namely, the anode and one or more screen grids. The division of the electron stream between a number of collecting electrodes gives rise to a noise known as *partition noise* in the anode circuit. Although the ratio in which the electron stream is divided between the screen grid and the anode is constant when electrons are counted over an appreciable period of time, the ratio inevitably varies if the count is made over a sufficiently small period of time; this argument is of the same nature as that used to account for shot noise. We can say that the division of electron beams introduces noise which is confined to tetrodes and more complex valves.

Because of the absence of partition noise, triodes are quieter than pentodes, and for this reason are frequently preferred in the early stages of receivers or amplifiers where it is essential to keep internally-generated noise to a minimum.

It will be noted that the expressions given above for the thermal agitation voltage and the shot noise current both gave the results in terms of the r.m.s. value. As a result of a number of tests which have been carried out it would appear that the r.m.s. value of noise gives a closer correlation to its annoyance value than the peak or mean value: in other words it is the *noise power* which is important.

Summarising the previous paragraphs we can say that the three principal sources of interference with which we are concerned are (1) from another channel, (2) from impulsive noise, and (3) from fluctuational noise. All three forms of interfering signal consist of a number of sinusoidal components, some of which fall within the passband of the receiver and cause interference by reacting with the components of the wanted signal. It is difficult to assess the effect of a large number of interference components reacting with a large number of components comprising the wanted signal, and we shall begin this investigation into interference by considering the reaction between a single wanted sinusoidal signal and a single unwanted interfering signal also of sinusoidal waveform. The result is afterwards extended to problems where many wanted and unwanted components are present: in this way it is possible to compare the performance of an f.m. receiver with an a.m. receiver when both are subjected to interference.

EQUIVALENT AMPLITUDE AND PHASE MODULATION

The effect of combining two components of different frequencies and different amplitudes can be determined by means of a vector diagram such as that shown in Fig. 3.1. The vector w represents the wanted carrier and rotates anticlockwise (the conventional positive direction) at one revolution per cycle of the carrier frequency f_w. The vector u represents the unwanted signal and also rotates anticlockwise at a velocity determined by its frequency f_u. To simplify the explanation we can assume the vector w to be stationary and that the vector u rotates with respect to it at the difference $(f_u - f_w)$ of the two frequencies. To find the resultant vector r we complete the triangle as indicated. As the vector u rotates, its arrow head traces a circle about the arrow head of the vector w to produce a resultant r which periodically varies in length

between the limits $(w - u)$ and $(w + u)$ and also varies in phase relative to the wanted carrier between the limits $+\phi$ and $-\phi$.

The variation in length of the resultant vector means that the resultant signal fed to the receiver is amplitude-undulated. This is illustrated in Fig. 3.2. The depth of the amplitude modulation depends on the ratio of u to w, and if the amplitude of the unwanted

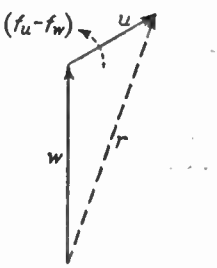


Fig. 3.1. Determination of the resultant vector r of two vectors w and u

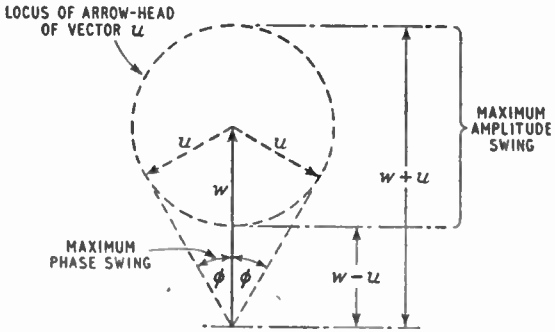


Fig. 3.2. Illustrating the resultant phase modulation and amplitude modulation due to two sinusoidal signals of different amplitudes and frequencies

signal is equal to that of the wanted signal, the amplitude modulation is 100 per cent and the resultant signal varies between a maximum and zero. It is one of the aims of f.m. receiver design to make the receiver unresponsive to amplitude variations of the input, and it is usually possible to achieve this provided the amplitude modulation is not too deep. We can therefore say that if the unwanted signal amplitude is small compared with that of the wanted signal the receiver will not respond to the amplitude modulation.

The variation in phase angle means that the resultant signal fed to the receiver is phase-modulated, and the receiver will naturally respond to this because there is no difference in nature between a frequency-modulated and phase-modulated signal. The peak value of the phase swing is $\pm\phi$, where

$$\phi = \sin^{-1}\frac{u}{w}$$

If $u/w = 0{\cdot}5$, $\sin\phi = 0{\cdot}5$ and $\phi = 30°$. We therefore obtain a peak phase swing which depends only on the relative amplitudes of the two carriers. It does not depend upon the frequency difference

$(f_w - f_u)$, and a given ratio of unwanted to wanted signal amplitude always produces the same peak phase swing.

If this phase swing is comparable with that of any phase modulation of the wanted carrier, then the unwanted signal will be audible and, in fact, the ratio of the two phase swings will give us the ratio of unwanted to wanted signal output.

If the unwanted signal amplitude is small compared with that of the wanted signal, the variations of phase angle ϕ as u rotates are almost sinusoidal. This variation is heard as a single beat note at the output of an f.m. receiver tuned to either of the two input signals. The variation of ϕ with time can be expressed

$$\phi = \phi_{pk} \sin \omega_d t$$

where ω_d is the angular difference frequency. Now angular frequency is the rate of change of phase and we thus have

$$\omega = \frac{d\phi}{dt}$$

$$= \frac{d}{dt}(\phi_{pk} \sin \omega_d t)$$

$$= \omega_d \phi_{pk} \cos \omega_d t$$

The peak value of the angular frequency swing ω_{max} occurs when $\cos \omega_d t$ is unity

$$\therefore \omega_{max} = \omega_d \phi_{pk}$$

from which it follows that

peak frequency swing = peak phase swing × frequency difference.

This, it must be emphasised, applies only when the unwanted signal amplitude is small compared with that of the wanted signal. When the two signals are of comparable amplitude this relationship is no longer true because the time variation of ϕ is no longer approximately sinusoidal. This also implies that the beat note heard at the output of the receiver is no longer a single frequency but has harmonics also.

INTERFERENCE FROM AN UNMODULATED CARRIER

Suppose the difference between the frequencies of the two input signals is 50 c/s. A 30° phase shift is equal to $\pi/6$ radians and is

equivalent to a peak frequency swing given by the following expression:

$$\text{peak frequency swing} = \text{peak phase swing} \times \text{difference frequency}$$

$$= \phi_{pk}(f_u - f_w)$$

$$= \frac{\pi}{6} \times 50 \text{ c/s}$$

$$= 26{\cdot}2 \text{ c/s}$$

In an f.m. system such as those used for broadcasting, in which the peak carrier frequency swing is 75 kc/s, a swing due to interference as small as 26 c/s is very small.

In fact, the ratio of unwanted to wanted signal at the f.m. detector output is given by

$$\frac{\text{unwanted signal output}}{\text{wanted signal output}}$$

$$= 20 \log_{10} \frac{\text{peak phase swing due to unwanted modulation}}{\text{peak phase swing due to wanted modulation}}$$

$$= 20 \log_{10} \frac{26}{75 \times 10^3}$$

$$= -20 \log_{10} (3 \times 10^3) \text{ dB approximately}$$

$$= -20 \times 3{\cdot}4771 \text{ dB}$$

$$= -70 \text{ dB nearly}$$

With such a large ratio of wanted to unwanted signal, the interference would be inaudible.

An a.m. receiver would respond to the equivalent amplitude modulation which is 50 per cent (for 30° phase swing) and the interfering signal is thus only 6 dB down on the peak value (100 per cent modulation) of the wanted signal. Such interference would be very obvious.

Thus, in this particular example, the f.m. receiver is greatly superior to the a.m. receiver. It can, in fact, be said that for small frequency differences such as 50 c/s the f.m. receiver is virtually free from interference.

The difference in performance is not so marked, however, when the frequency difference is greater. At 15 kc/s, for example, which is commonly taken as the upper frequency limit in high-fidelity

reproduction, the phase difference of $\pi/6$ radians corresponds to a peak frequency swing given by

$$\begin{aligned}\text{peak frequency swing} &= \text{peak phase swing} \times \text{frequency difference}\\ &= \phi_{pk}(f_u - f_w)\\ &= \frac{\pi}{6} \times 15 \text{ kc/s}\\ &= 7{\cdot}86 \text{ kc/s}\end{aligned}$$

This is comparable with a peak frequency swing of 75 kc/s. In fact the difference in output due to unwanted and wanted signals at the f.m. detector output is given by

$$\begin{aligned}&20 \log_{10} \frac{\text{peak phase swing due to unwanted modulation}}{\text{peak phase swing due to wanted modulation}}\\ &= 20 \log_{10} \frac{7{\cdot}86}{75} \text{ dB}\\ &= -20 \log_{10} 10 \text{ dB approximately}\\ &= -20 \text{ dB}\end{aligned}$$

For an a.m. receiver the change in frequency difference from 50 c/s to 15 kc/s has no effect on the level of interference: the unwanted signal is still −6 dB relative to the wanted signal. Thus at 15 kc/s the f.m. receiver gives 14 dB more protection against interference due to a single unwanted sinusoidal signal than an a.m. receiver.

If the frequency difference is increased above 15 kc/s (to 75 kc/s for example) the f.m. receiver is inferior to the a.m. receiver in that a larger ratio of unwanted to wanted output is obtained at the detector. However, this is of no consequence because the a.f. amplifier or the ear of the listener attenuates rapidly above 15 kc/s.

These values were evaluated for a ratio of wanted to unwanted signal amplitude of 2 : 1; for smaller values of unwanted signal amplitude the superiority of the f.m. receiver is even more marked and, in fact, the frequency of a small amplitude interfering signal can be swung through that of the signal to which an f.m. receiver is tuned without trace of the audible beat note which is heard in a.m. receivers.

If in Fig. 3.2 the amplitude of the interfering signal is small compared with that of the wanted signal, the angle ϕ is small and

$\sin\phi$ is approximately equal to ϕ provided this is expressed in radians. Thus the peak phase swing ϕ_{pk} is approximately equal to the ratio of the unwanted signal amplitude to that of the wanted signal. If

e_u = interfering-signal carrier amplitude

and

e_w = wanted-signal carrier amplitude

we have

$$\phi_{pk} = \frac{e_u}{e_w}$$

and the peak frequency swing due to the interfering signal is given by

$$\begin{aligned}\text{peak frequency swing} &= \text{peak phase swing} \times \text{frequency difference} \\ &= \phi_{pk} \times (f_u - f_w) \\ &= \frac{e_u(f_u - f_w)}{e_w}\end{aligned}$$

For 100 per cent modulation the peak frequency swing is equal to the deviation f_d

$$\therefore \frac{\text{output of detector due to interfering signal}}{\text{output of detector for 100 per cent modulation}}$$

$$= \frac{e_u}{e_w} \cdot \frac{(f_u - f_w)}{f_d}$$

An a.m. receiver gives an output at the detector proportional to the input carrier amplitude. Thus

$$\frac{\text{output of detector due to interfering signal}}{\text{output of detector due to wanted signal}} = \frac{e_u}{e_w}$$

If the wanted signal is taken as representing 100 per cent modulation, we can compare this expression directly with that given above for an f.m. receiver. The comparison shows that the output due to the interfering signal in an f.m. receiver is less than that for an a.m. receiver in the ratio $(f_u - f_w)/f_d$. The greatest value that $(f_u - f_w)$ can have is 15 kc/s because higher frequencies are inaudible. Thus the factor of improvement of f.m. over a.m. for 75 kc/s deviation is 75/15, i.e. 5 : 1 at 15 kc/s and is greater at lower frequencies. This represents an improvement of at least 14 dB.

The above expressions show that for a given frequency difference, interference in an f.m. receiver can be decreased by increasing the deviation frequency but this increases the technical difficulties of f.m. transmitter and receiver design and increases the bandwidth occupied by an f.m. transmission, reducing the number of transmissions which can be accommodated in a given waveband. 75 kc/s is the greatest deviation normally employed in broadcasting.

TRIANGULAR NOISE SPECTRUM

We have shown that the output at the detector of an f.m. receiver for the simultaneous reception of two sinusoidal signals is proportional to the difference between their frequencies, i.e. is linearly

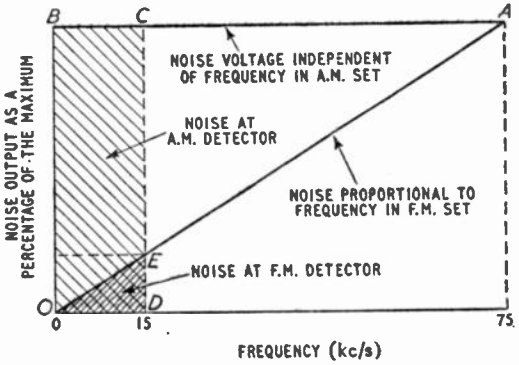


Fig. 3.3. Illustrating the relative noise outputs at the detectors of an a.m. and f.m. receiver

related to the frequency of the detector output. This is illustrated in Fig. 3.3 by the straight line OA which passes through the origin. This confirms a point made earlier, namely, that for interfering signals which produce low audio outputs at the detector, an f.m. receiver gives less audio output than an a.m. receiver. For an a.m. receiver the output due to an interfering signal depends only on its magnitude relative to that of the wanted signal and not on its frequency. This is represented by the straight line BC parallel to the frequency axis.

The shaded area ODE represents the noise output from an f.m. receiver: this area does not include frequencies in excess of 15 kc/s because they are inaudible. The shaded area OBCD represents the noise output from an a.m. receiver. If we assume that the

maximum noise outputs of both receivers are equal, a comparison of the two shaded areas gives an impression of the superiority of the f.m. receiver over the a.m. receiver.

INTERFERENCE FROM A FREQUENCY-MODULATED CARRIER

We have so far discussed the behaviour of f.m. and a.m. receivers to two unmodulated sinusoidal signals, one of which was assumed to be wanted and the other unwanted. We shall now consider what happens when one or both of the signals is modulated.

Suppose an f.m. receiver is tuned to a modulated signal and that an unmodulated signal also falls within the bandwidth. If the difference in carrier frequencies is less than 15 kc/s, there will be an audio output due to interference between these carriers at a single frequency if the amplitude of the unwanted signal is small compared with that of the wanted signal, but accompanied by harmonics if the carriers have comparable amplitudes. The amplitude of the interference signal at the detector output is proportional to its frequency and, provided the unwanted carrier amplitude is small compared with that of the wanted carrier, interference is not likely to be serious. If the frequency difference is greater than 15 kc/s, say 30 kc/s for example, beat notes from the detector are not audible and it would appear that no interference would be heard. In practice there is interference which is heard as a swish each time the frequency of the wanted signal sweeps through that of the unwanted signal.

If the unwanted carrier is frequency-modulated, interference can take two forms. Firstly, the modulation of the unwanted carrier may be audible: secondly, the swish due to one carrier passing through the other may be heard. If the two transmissions have the same carrier frequency and are of equal amplitude, both programmes are heard equally well. If now the amplitude of the unwanted carrier is slowly decreased, the unwanted programme becomes rapidly weaker and, under favourable conditions, may be inaudible when the unwanted carrier amplitude is less than one half that of the wanted signal. Thus, as a broad generalisation we may say that in co-channel interference in f.m. reception the two programmes are either heard at equal strength or else one only is heard.

The monopolisation of an f.m. receiver by a signal whose amplitude only slightly exceeds that of other signals on the same or

substantially the same carrier frequency is known as the *capture effect*. The *capture ratio* is the ratio, usually expressed in dB, of the amplitudes of two signals on the same frequency when the modulation of one is just audible relative to that of the other. The capture ratio of an f.m. receiver depends on its insensitivity to amplitude-modulated signals and to achieve a good, i.e. a small capture ratio the receiver must be designed to be as insensitive to a.m. signals as possible. The behaviour of an f.m. receiver towards a.m. signals is usually expressed in terms of the *a.m. suppression ratio* which is defined as the ratio of the output due to a frequency-modulated signal (modulated to ± 30 kc/s, equivalent to 40 per cent modulation) to the output due to amplitude modulation (modulated to a depth of 40 per cent also) when both carriers are of equal amplitude.

To facilitate measurement a single carrier may be used which is frequency-modulated at say 100 c/s and amplitude-modulated at say 2,000 c/s. A well-designed f.m. receiver may have an a.m. suppression ratio of 50 dB giving a capture ratio of say 4 dB: this means that an interfering signal of more than half the amplitude of the wanted signal is inaudible. A typical commercial receiver may have an a.m. suppression ratio of 30 dB: for this the corresponding capture ratio is 12 dB, implying that the interfering signal must have an amplitude one quarter that of the wanted signal for its modulation to be inaudible. A receiver with an a.m. suppression ratio of less than 30 dB would have a capture ratio greater than 12 dB: this represents rather a poor performance. A.m. suppression ratios should exceed 30 dB and this needs careful design of the limiter and/or detector stages of the receiver.

If the unwanted carrier amplitude is, say, one tenth that of the wanted signal the level of interference at the detector output may be as low as -55 dB relative to that of the wanted programme. At such a low level the programme of the unwanted signal is completely inaudible but interference is still present in the form of the swishing noise due to interaction between the carriers. This type of noise is particularly distracting and can mar the enjoyment of a musical programme even at a level as low as -55 dB.

MULTI-PATH DISTORTION

Signals from an f.m. transmitter sometimes arrive at a receiving aerial by a number of paths. The direct path is along the straight line connecting the transmitter to the receiving aerial but signals may take other, longer, paths which include one or more reflections

at large objects such as hills or gasholders. When such paths exist, two or more signals arrive simultaneously at the receiver. All the signals have the same modulation but the differing path lengths cause the programmes on all signals other than the direct one to be delayed in time relative to the programme which is on the direct signal.

At any given instant therefore the various signals are at different points in the modulation waveform. Interference can thus result as in co-channel interference.

For path differences of less than 5 miles the effects of multi-path reception are generally negligible because the time delay in the modulation waveform is too small to be detected. For path differences between 5 and 20 miles the interference can result in serious distortion which is particularly noticeable in the reproduction of the piano.

In severe examples of multi-path interference the distortion can be so appalling that the listener may be convinced that there is something loose in the loudspeaker! For path differences exceeding 20 miles the swishing noise previously referred to becomes troublesome and the distortion is so severe that the programme is scarcely recognisable.

The distortion which accompanies multi-path reception can be reduced by improving the a.m. suppression ratio of the receiver. A receiver with 35 dB a.m. suppression may give negligible distortion on a signal which is badly distorted on another receiver having only 20 dB suppression. In general, at least 35 dB a.m. suppression is needed to eliminate the worst effects of multi-path distortion, and this calls for even greater care in design than for the 30 dB needed to give a worthwhile capture ratio.

INTERFERENCE FROM WHITE NOISE

The white noise arising from the aerial circuit and the early stages of an f.m. receiver has a continuous spectrum which is spread uniformly over the passband of the receiver. In general the noise voltages are small compared with likely values of wanted-carrier amplitude, and we can therefore assess the effect of each noise-frequency component within the receiver passband by the methods employed above. We can assume that the amplitude modulation resulting from the addition of the wanted and unwanted signals will yield no audible result at the f.m. detector output but that the phase modulation will. The magnitude of the noise at the f.m.

detector output compared with that due to a 100 per cent modulated wanted signal is given by

$$\frac{\text{output of detector due to noise signal}}{\text{output of detector for 100 per cent modulation at the wanted-signal frequency}} = \frac{e_u}{e_w} \cdot \frac{f_u - f_w}{f_d}$$

where e_u = noise-signal amplitude

e_w = wanted-signal carrier amplitude

f_u = noise-signal frequency

f_w = wanted-signal frequency

f_d = deviation frequency

This expression is illustrated in Fig. 3.3, the so-called triangular noise spectrum.

In this diagram it is shown that the noise output from an f.m. receiver is only 1/5th that from an a.m. receiver for a given ratio of noise to wanted signal amplitude. This advantage of the f.m. receiver is due to the wider passband of the receiver: only those noise components which are within 15 kc/s of the wanted carrier frequency can produce audible output at the detector. But sidebands of the wanted signal up to 75 kc/s of the carrier can produce wanted audio output.

In addition to this basic improvement (which applies equally to co-channel interference) there is a further improvement because the noise output from an f.m. detector is directly proportional to frequency.

In order to assess this particular advantage we must remember that it is the noise power which is important, and to compare the noise due to an f.m. receiver (which has a triangular spectrum) with that due to an a.m. receiver (which has a rectangular spectrum) we must compare the power contributed by all the noise components between zero frequency and the upper limit of audibility. As the noise has a continuous spectrum, this can be achieved by means of integration.

Consider first the a.m. receiver and let us suppose that the detector output due to noise has an r.m.s. voltage of e which is generated across a circuit of resistance R. The noise power at a single frequency is given by e^2/R, and if the noise voltage has a constant value independent of frequency, the total noise power

contributed by all the noise components between zero frequency and the upper limit of audibility f_0 is given by

$$\int_0^{f_0} \frac{e^2}{R}\,df = \frac{e^2}{R}\int_0^{f_0} df$$

$$= \frac{e^2 f_0}{R}$$

Consider now the f.m. receiver. For the same amplitudes of wanted and noise signals the noise voltage also has an r.m.s. value of e at the frequency f_0. Because the spectrum is triangular, the noise voltage at any other frequency f is given by ef/f_0. The noise power at this particular frequency is thus $e^2f^2/f_0{}^2R$ and the total noise power contributed by all the components between zero and the upper limit f_0 is given by

$$\int_0^{f_0} \frac{e^2 f^2}{f_0{}^2 R}\,.\,df = \frac{e^2}{f_0{}^2 R}\int_0^{f_0} f^2\,.\,df$$

$$= \frac{e^2 f_0{}^3}{3f_0{}^2 R}$$

$$= \frac{e^2 f_0}{3R}$$

Comparing the two results we have

$$\frac{\text{noise for a triangular spectrum}}{\text{noise for a rectangular spectrum}} = \frac{e^2 f_0}{3R}\,.\,\frac{R}{e^2 f_0}$$

$$= \tfrac{1}{3}$$

The noise power in the f.m. receiver is only one third that due to the a.m. receiver, an improvement of 4·8 dB to add to the 14 dB already deduced.

INTERFERENCE FROM IMPULSIVE NOISE

Noise Peak small compared with Wanted Carrier Amplitude

When a steep-sided signal such as a pulse is applied to an f.m. receiver it gives rise to a damped wavetrain of the form illustrated in Fig. 3.4. This signal is at the centre frequency of the i.f. pass-band and its amplitude rises and falls with time. To assess the

effect of such an interfering signal, when received at the same time as the wanted signal, we will first assume that the unmodulated-signal amplitude is at all times small compared with that of the carrier of the wanted signal. We will further assume that the receiver is accurately tuned to the wanted signal: the carrier is also at the mid-band frequency of the i.f. amplifier and therefore has the same frequency as the damped oscillation due to the noise pulse.

There is, therefore, a fixed phase relationship between the wanted carrier and the unwanted noise signal. This is illustrated in Fig. 3.5(a) in which the vector w represents the wanted signal and vector u the unwanted signal. This diagram shows u as very small compared with w; this may be taken as representing conditions

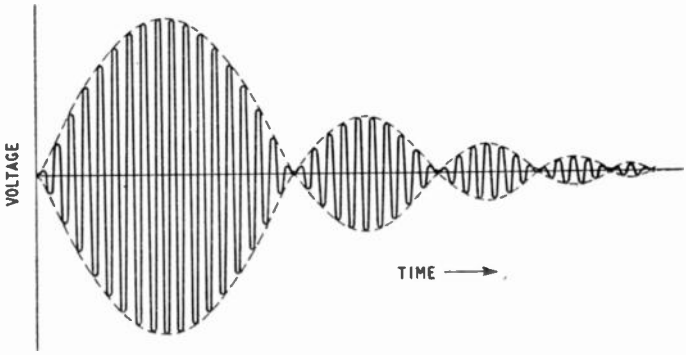


Fig. 3.4. Form of damped wave induced in a tuned amplifier by a steep-sided pulse

near the beginning of the damped wavetrain and before the first maximum is reached.

The resultant vector r makes an angle ϕ with the carrier, such that

$$\tan \phi = \frac{u \sin \theta}{w + u \cos \theta}$$

where θ is the phase angle between the wanted and unwanted signals.

Now consider conditions a little later in time when the damped wavetrain has reached its first maximum. The vector u is now

longer than it was, although still shorter than w: the phase angle θ is unchanged, both signals still being on the same frequency. The vector diagram now has the form shown in Fig. 3.5(b).

Comparing this with vector diagram (a) we can see that the resultant vector r has increased in length and that the phase angle ϕ has also increased.

If now we imagine that the length of u varies in time in accordance with the amplitude of the damped wavetrain we can see that the

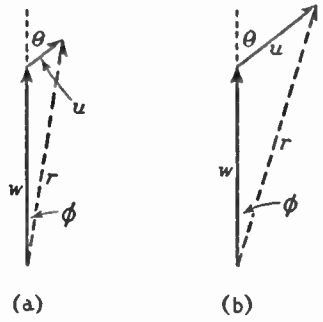


Fig. 3.5. Illustrating the variation in length of r and phase angle ϕ as u varies in length, when the phase angle θ remains constant and the noise vector is small compared with the wanted carrier

resultant vector varies in length, and the phase angle ϕ varies in value, in much the same way as u. The variations in length of r constitute amplitude modulation, and in a well-designed f.m. receiver will produce no output at the detector. The variations in phase angle ϕ constitute phase modulation and will produce an audio output, the nature of which we shall now consider.

From the equation above it can be seen that if u is small compared with w

$$\tan \phi \simeq \frac{u}{w} \sin \theta$$

and if θ and ϕ are both small

$$\phi \simeq \frac{u}{w} . \theta$$

Thus the variation with time of the phase angle of the resultant r with respect to the carrier is of the same waveform as the envelope of the damped oscillation induced in the i.f. amplifier by the noise

pulse. This waveform is illustrated in Fig. 3.6(a). The output from the f.m. detector is proportional to the differential with respect to the time of this and therefore has a waveform of the type shown in Fig. 3.6(b).

This also has an appearance similar to that of a damped wavetrain. The individual cycles of this wave are not audible because their frequency is supersonic, and whether they are heard at all depends on the net area either above or below the axis. If the area under the curve but above the time axis is equal to that above the curve but below the axis, the waveform would be inaudible.

In practice there is usually a resultant area, which means that the detector output is heard. The sound depends on the distribution of energy with frequency in the wave, and for a waveform of this type the energy tends to be distributed in direct proportion to frequency: in other words the spectrum is triangular. As a result of the concentration of energy into the upper frequencies, the waveform has a high-pitched sound usually described as a " click ". Such

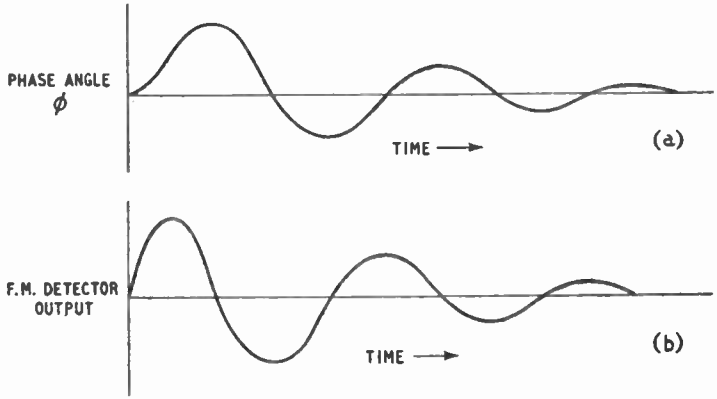


Fig. 3.6. The variation of phase angle with time for the vector diagram of Fig. 3.5 is illustrated at (a) and the corresponding f.m. detector output is illustrated at (b)

" clicks " are not very distracting and do not constitute serious interference with the wanted programme.

It was assumed above that there was a fixed phase relationship between the damped oscillation produced by the noise pulse and the wanted carrier. It is unlikely in practice that an f.m. receiver will be so perfectly tuned, and there will be a difference in frequency between the oscillation and the carrier. This can be represented

in the vector diagram by assuming the vector w, representing the wanted carrier, to be stationary and that the smaller vector u, representing the damped oscillation, rotates about the arrowhead of the carrier vector at the difference frequency. At the same time as the vector u rotates, it also varies in length as pictured in Fig. 3.4 on page 40.

Thus, the arrowhead of the noise vector traces out a locus such as that shown dotted in Fig. 3.7. The phase angle between the resultant vector and the carrier vector could be determined for each position of the noise vector, and the phase angle-time curve could thus be determined. Differentiation of this gives the waveform of the detector output. If this is repeated for a number of values of

Fig. 3.7 (right). *Typical locus of noise vector when the noise signal and the carrier have different frequencies, the noise vector being small compared with the wanted carrier amplitude*

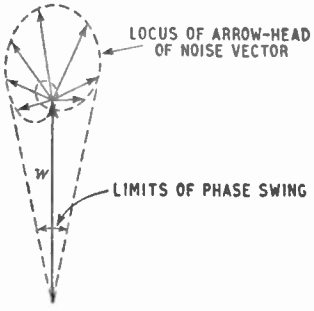
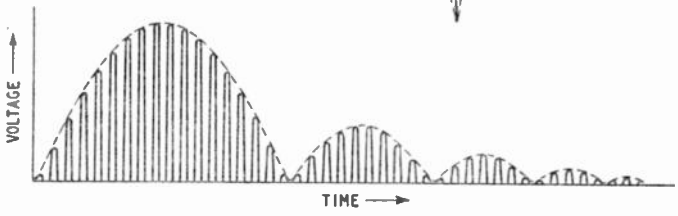


Fig. 3.8 (below). *Damped wave of Fig. 3.4 when negative-going signals are removed by an a.m. detector*



difference frequency and for a number of different initial phase angles θ, it will be found that the detector output always has a waveform of the same nature as that shown in Fig. 3.6 and has the audible sound of a " click ".

This should be compared with the sound likely to be heard when a noise pulse is picked up by an a.m. receiver. It results in a damped oscillation in the i.f. amplifier, the waveform being similar to that shown in Fig. 3.4. This is now applied to an a.m. detector which has the effect of eliminating either the positive-going or the negative-going excursions of the waveform, leaving a unidirectional series of half-cycles at the i.f. frequency as shown in Fig. 3.8. These

are integrated by the detector to give an audio output of the form shown in Fig. 3.9. This is completely different from the f.m. detector (Fig. 3.6[a]). It has a large energy content since there is no area beneath the axis, and the distribution of energy tends to be uniform throughout the spectrum. There is thus considerable low-frequency content and the detector output tends to have the sound of a " pop ", which is more distracting and constitutes more serious interference than a " click ".

In conclusion we may say that impulsive noise whose peak value is less than the amplitude of the wanted carrier produces *clicks* in

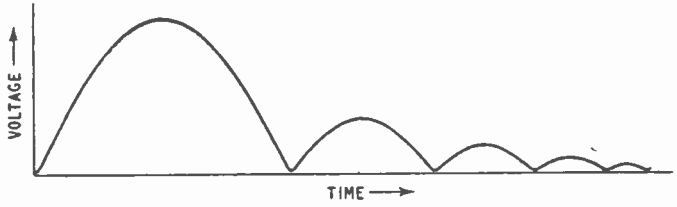


Fig. 3.9. Output of a.m. detector when the input is the damped wave of Fig. 3.4

an f.m. receiver and *pops* in an a.m. receiver. The ratio of unwanted noise output to wanted signal output is of the order of 20 dB better for an f.m. than for an a.m. receiver, the factor of improvement being approximately equal to that which applies to white noise. The audible improvement is better than the figure of 20 dB suggests, because the noise in the f.m. receiver is high-pitched in character.

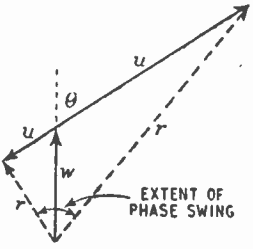
Noise Peak large compared with Wanted Carrier Amplitude

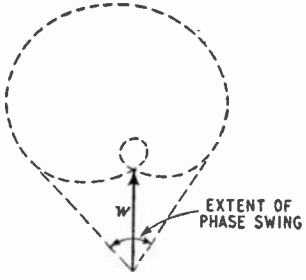
Let us now consider the interference due to a pulse of noise the peak value of which is greater than the amplitude of the wanted carrier. The damped oscillation is induced in the i.f. amplifier in the same way as for a small noise pulse. To simplify the problem, suppose that the receiver is accurately tuned to the carrier: there is then a fixed phase relationship between the vector w representing the wanted carrier and the vector u representing the damped oscillation. The length of u varies and even reverses as the oscillation dies away, but the angle θ remains fixed. The resultant vector r varies in length to give amplitude modulation which is ignored by the f.m. receiver, but the angle ϕ between the resultant vector r and the carrier w varies to give phase modulation. The total extent of the swing of ϕ is indicated in Fig. 3.10, and for such a

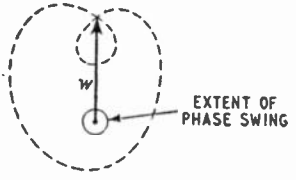
small variation the output of the f.m. detector has the same form as described previously when the peak noise signal is small. In other words, a " click " is heard for each noise pulse received.

In practice the receiver is not likely to be so perfectly tuned that θ remains constant. A frequency difference (if only a small one) is

Fig. 3.10 (right). *Illustrating the extent of the phase variation ϕ as u varies in length when the phase angle θ remains constant and the noise vector is large compared with the wanted carrier*







(a) (b)

Fig. 3.11. Typical loci of noise vector when the noise vector and carrier have different frequencies, the noise vector being large compared with the wanted carrier amplitude

inevitable between the carrier and the damped oscillation. This can be allowed for if we assume the vector w to be fixed and the vector u to rotate about its arrowhead at the difference frequency. As u rotates, it also varies in length in accordance with the damped oscillation of Fig. 3.4. The locus of the arrowhead of the vector u can trace out a number of different-shaped figures depending on the frequency difference and on the initial phase angle at the instant of receipt of the noise pulse.

Two particular figures are significant and are illustrated in Fig. 3.11. In (a) the figure is such that the phase cycle ϕ executes variations to either side of the carrier, none of which are particularly large; they do not exceed 90° for instance. For such an example the variation of ϕ with time is again similar to that described above, and the result is again that a " click " is produced at the

detector output. The rotation of the vector u has had no significant effect on the nature of the interference. In diagram (b) the loop traced out by the arrowhead of vector u embraces the origin of the diagram, and if the resultant vector r is drawn for various positions of vector u it will be found that the phase angle ϕ has either advanced or has retreated through an angle of 360° during the damped oscillation.

The phase angle does not return to zero as in earlier examples but stays at +360° or −360°, the variation of ϕ with time having the form shown in Fig. 3.12(a). To find the waveform of the f.m.

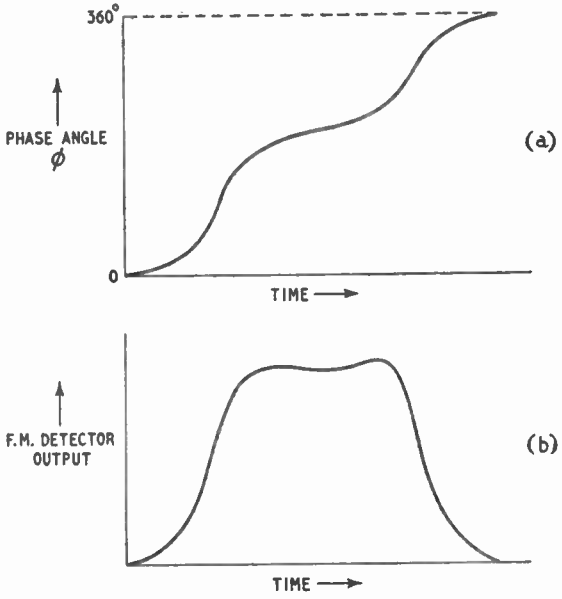


Fig. 3.12. A typical variation of phase angle with time for the vector diagram of Fig. 3.11 (b) is shown at (a) and the corresponding f.m. detector output at (b)

detector output we must differentiate this curve; we then obtain a waveform of the type shown in Fig. 3.12(b). Here the waveform is shown as a double-humped curve but it can take the form of two separate pulses. Such waveforms are similar to those obtained at the detector of an a.m. receiver subjected to interference and have the audible sound of a " pop " which is more disturbing than a " click " and constitutes more severe interference. " Pops " can

occur only when the peak noise exceeds the carrier amplitude but not all the noise pulses are heard as " pops "; a proportion are heard as " clicks " as shown in Fig. 3.11(a).

Threshold of Improvement

If an f.m. receiver is subjected to impulsive interference which is initially small compared with the wanted carrier amplitude but is steadily increased until it is large compared with the carrier amplitude, the interference is first heard as a series of faint " clicks " but is later heard as a mixture of " clicks " and " pops ". The onset of the " pops " constitutes a deterioration in signal-noise ratio which occurs, as can be seen from Fig. 3.11 when the peak noise in the i.f. amplifier is equal to the wanted carrier amplitude.

This level of noise therefore marks a threshold. For noise inputs below the threshold the signal-noise ratio is in general acceptable; for noise levels above it, the signal-noise ratio is much inferior and may not be acceptable. At the threshold there is an abrupt change in signal-noise ratio for a small change in the value of the peak noise. This effect has something in common with the capture effect described earlier.

PRE-EMPHASIS AND DE-EMPHASIS

We have shown that, when interference is experienced in f.m. reception, the noise output from the detector tends to be proportional to the frequency and is hence concentrated in the upper half of the audio spectrum. This noise can be made less annoying by including, after the detector, a network which attentuates high frequencies relative to middle and low audio frequencies. This frequency discrimination also applies, of course, to the wanted programme but the balance of high and low frequencies can be restored and an overall flat response obtained by using in the modulator circuits of the transmitter a network with a complementary frequency characteristic, i.e. one which boosts the high frequencies relative to the middle and low ones. The high-frequency boost is known as *pre-emphasis* and the corresponding loss as *de-emphasis*; this technique is also used in disk recording to reduce the effects of surface noise during reproduction.

It was at first thought that pre-emphasis and de-emphasis in f.m. broadcasting would give a considerable improvement in signal/noise ratio and a substantial lift of high frequencies was used. The networks had a time constant of 100 μsec; as shown in the

appendix this corresponds to a lift of 19·5 dB at 15,000 c/s, an equal loss being introduced, of course, in the a.f. section of the receiver. It was found, however, that the level of middle and low-frequency modulating signals at the transmitter had to be so reduced to avoid over-modulation and consequent distortion of high-frequency signals that the improvement in signal-noise ratio was largely lost. The time constant was then reduced to 75 μsec, and when the B.B.C. introduced its f.m. service, the time constant was made even less, 50 μsec, corresponding to a 13·6 dB lift at 15,000 c/s. Measurements indicate that on impulsive interference the improvement in signal/noise ratio due to the use of a de-emphasis network does not exceed 2 dB.

APPENDIX

RELATIONSHIP BETWEEN TIME CONSTANT AND FREQUENCY RESPONSE

A simple RC circuit such as that illustrated in Fig. 3.13 has a response which falls as frequency rises: it is, in fact, commonly used for de-emphasis purposes in f.m. receivers. The frequency response of the network can be calculated in the following way. The current in the circuit is given by

$$i = \frac{V_{in}}{R + 1/j\omega C}$$

and the output voltage is given by

$$V_{out} = i/j\omega C$$

$$\therefore \frac{V_{out}}{V_{in}} = \frac{1/j\omega C}{R + 1/j\omega C}$$

$$= \frac{1}{1 + j\omega CR}$$

This demonstrates that the frequency response of the network depends only on the product CR and not on the individual values

of resistance and capacitance. The response is therefore determined by the time constant $T(=CR)$ according to the expression

$$\frac{V_{out}}{V_{in}} = \frac{1}{1 + j\omega T}$$

from which

$$\left|\frac{V_{out}}{V_{in}}\right| = \frac{1}{\sqrt{(1 + \omega^2 T^2)}}$$

The frequency response in decibels is given by

$$20 \log_{10}\left|\frac{V_{out}}{V_{in}}\right| = 20 \log_{10} \frac{1}{\sqrt{(1 + \omega^2 T^2)}}$$

$$= -20 \log_{10} \sqrt{(1 + \omega^2 T^2)}$$

$$= -10 \log_{10} (1 + \omega^2 T^2)$$

This expression also applies to the rising response obtained if the output is taken from the resistor in Fig. 3.13. Furthermore, the

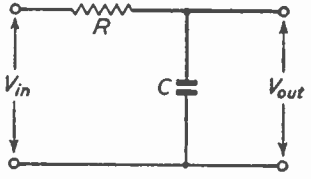


Fig. 3.13. Circuit commonly used for de-emphasis in f.m. receivers

response obtained from a simple network of resistance and inductance also obeys the same expression provided the time constant is taken as L/R.

This expression is hence of universal application and gives the frequency response of any simple RC or RL network.

As examples of the use of this expression we will determine the top lift or top cut given at 15,000 c/s by a network of 50 μsec time constant. Substituting in the expression we have

$$\text{Response} = -10 \log_{10} (1 + 6{\cdot}284^2 \times 15{,}000^2 \times 50^2 \times 10^{-12}) \text{ dB}$$

$$= -10 \log_{10} (1 + 22) \text{ dB}$$

$$= -10 \log_{10} 23 \text{ dB}$$

$$= -13{\cdot}6 \text{ dB}$$

This is the response of a network with a time constant of 50 μsec. An RC network with such a time constant for which the resistance is 100,000 ohms must have a capacitance given by

$$RC = 50 \times 10^{-6} \text{ sec}$$

$$C = \frac{50 \times 10^{-6}}{100{,}000} \text{ F}$$

$$= 500 \times 10^{-12} \text{ F}$$

$$= 500 \text{ pF}$$

Suppose we wish to know what value of time constant gives a relative boost of 10 dB at 15 kc/s. Substituting in the above expression we have

$$\text{response} = -10 \log_{10} (1 + \omega^2 T^2)$$

$$-10 = -10 \log_{10} (1 + 6{\cdot}284^2 \times 15^2 \times 10^6 \times T^2)$$

$$\therefore 1 + 6{\cdot}284^2 \times 15^2 \times 10^6 \times T^2 = 10$$

$$T^2 = \frac{9}{6{\cdot}284^2 \times 15^2 \times 10^6}$$

$$\therefore T = \frac{3}{6{\cdot}284 \times 15 \times 10^3}$$

$$= 32{\cdot}6 \ \mu\text{sec}$$

4

GENERATION OF FREQUENCY-MODULATED WAVES

Introduction

WE shall now consider some of the methods which can be used to produce a frequency-modulated wave. Such methods have a number of applications of which the most obvious is in f.m. transmitters. The frequency modulator is the most important stage in an f.m. transmitter and, if the transmitter is used for high-quality sound broadcasting, the modulator must be linear to a very high degree: that is to say, the frequency displacement of the modulated carrier output must be strictly proportional to the amplitude of the modulating signal input. It is not by any means easy to design modulators to satisfy such stringent requirements.

Frequency modulators are also extensively used in electronic test gear such as the sweep generators used in conjunction with oscilloscopes to display the frequency response of radio receivers, and here the linearity requirements are not quite so difficult to satisfy.

VARIABLE-CAPACITANCE FREQUENCY MODULATORS

The resonance frequency of an LC circuit is given approximately by the expression

$$f = \frac{1}{2\pi\sqrt{(LC)}}$$

and it follows that the frequency can be varied by varying either the inductance L or the capacitance C. The required variations in frequency are usually small compared with the average frequency and only a small change in L or C is necessary. Suppose frequency modulation is achieved by variation of capacitance and that the

frequency changes by Δf when the capacitance changes by ΔC. We then have

$$f = \frac{1}{2\pi\sqrt{(LC)}} \qquad \ldots\ldots(1)$$

and

$$f + \Delta f = \frac{1}{2\pi\sqrt{L(C + \Delta C)}} \qquad \ldots\ldots(2)$$

Dividing equation (2) by (1) we have

$$1 + \frac{\Delta f}{f} = \sqrt{\left(\frac{C}{C + \Delta C}\right)}$$

$$= \frac{1}{\sqrt{(1 + \Delta C/C)}}$$

If ΔC is small compared with C

$$\frac{1}{\sqrt{(1 + \Delta C/C)}} \simeq \frac{1}{1 + \Delta C/2C}$$

$$\simeq 1 - \frac{\Delta C}{2C}$$

$$\therefore \frac{\Delta f}{f} = -\frac{\Delta C}{2C}$$

The fractional change in frequency is thus approximately half the fractional change in capacitance. The minus sign is a reminder that capacitance must be increased to decrease frequency and decreased to increase it. As a numerical example, in an oscillator operating at 10 Mc/s and tuned by a 50-pF capacitor, a 50 kc/s change in frequency requires a capacitance change given by

$$\Delta C = \frac{2\Delta f C}{f}$$

The minus sign is omitted because we are not asked whether the capacitance should be increased or reduced but merely what the change should be. Substituting the numerical values

$$\Delta C = \frac{2 \times 50 \times 10^3 \times 50}{10 \times 10^6} \text{ pF}$$

$$= 0{\cdot}5 \text{ pF}$$

One practical modulator of this type is the mechanical system used in some sweep generators. The modulator consists of a moving-coil movement similar to that used in loudspeakers but carrying instead of a diaphragm one plate of a parallel-plate capacitor, the other plate being fixed. If this capacitor is included in an oscillator circuit, any movement of the moving-coil causes a change of capacitance and hence a change in oscillator frequency. If the moving-coil is supplied with current of sawtooth waveform, the movement of the coil is such as to produce a frequency-modulated output in which frequency varies linearly over the range of operation and then returns rapidly to the opposite extreme to begin the next linear sweep.

VARIABLE-INDUCTANCE FREQUENCY MODULATORS

If the above calculation is repeated for an LC circuit containing a fixed capacitance, it will be found that variations of inductance also produce frequency changes, the fractional change being approximately half the fractional change of inductance. Successful frequency modulators operating on this variable-inductance

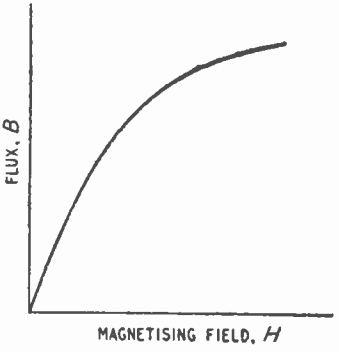


Fig. 4.1. Form of B-H curve for magnetic material such as ferrite

principle can also be constructed, and have been used in sweep generators. The principle of one such frequency modulator can be approached in the following way.

Suppose the inductor of an r.f. oscillator has a core of permeable material such as ferrite. The inductance of the coil depends on the permeability of the core, and this in turn depends on the state of magnetisation of the core and to some extent on the amplitude of r.f. current in the winding. The B–H curve for the core material may have a shape similar to that shown in Fig. 4.1, and if the r.f.

current is small the effective permeability is given by the slope of the B–H curve at the origin. If now the operating point is moved along the curve by placing a coil carrying direct current around the r.f. inductor, the slope of the curve is smaller, i.e. the incremental permeability is less, causing a reduction in inductance and an increase in oscillator frequency.

In this way it is possible to control the frequency of an oscillator by adjustment of a direct current. If an alternating (modulating)

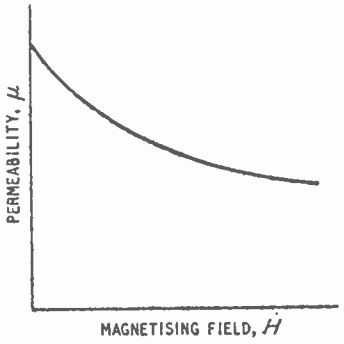


Fig. 4.2. Form of μ-H curve corresponding to Fig. 4.1

current is superimposed on the direct current, a frequency-modulated oscillation results. The variation of incremental permeability with H, and hence with direct current amplitude, is illustrated in Fig. 4.2: the curve has no linear portion but, provided the modulating current is kept small compared with the direct current, the departure from linearity will be found to be adequate for many applications.

Variable permeability frequency modulators of this type have been used for automatic frequency correction (a.f.c.) in radio receivers. In this application the r.f. winding is included in the oscillator circuit of the frequency changer and the modulator winding is supplied with current from a valve which is connected to a discriminator at the output of the i.f. amplifier. If the receiver is accurately tuned there is no output from the discriminator, but any mistuning of the oscillator gives a discriminator output which varies the valve current and alters the permeability of the core so so as to offset the initial mistuning. It is not possible by such a circuit to neutralise mistuning completely, but the departure from correct tuning can be reduced by a factor of the order of 5 : 1.

The principle of one simple form of frequency modulator is illustrated in Fig. 4.3(a). This shows a capacitive microphone connected across the frequency-determining LC circuit of an oscillator. Sound waves striking the diaphragm of the microphone vary its capacitance and change the oscillator frequency to give a frequency-modulated output. Such a simple frequency-modulating system is hardly practicable because the changes in capacitance are usually too small to give worthwhile results. With slight elaboration, however, this circuit can be made good enough for use in communications systems. As an example Fig. 4.3(b) gives in simplified form a circuit of the modulator of a walkie-talkie equipment.

This differs from Fig. 4.3(a) in that the microphone capacitance C_1 is tuned by an inductor L_1 to a frequency approximately equal to that of the oscillator. By this means the variations in reactance of

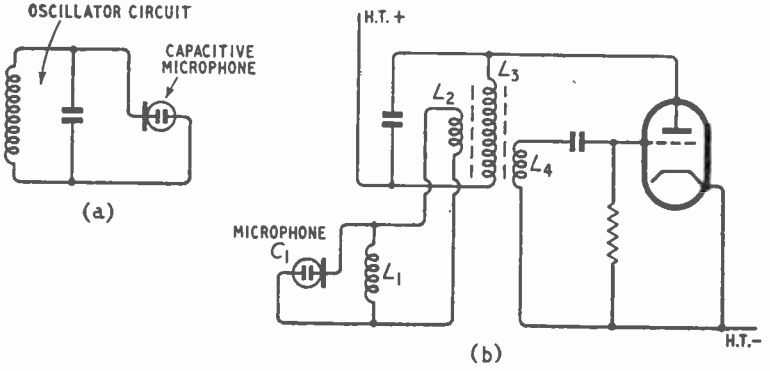


Fig. 4.3. Basic circuit of capacitive microphone modulator (a) and simplified modulator circuit of a walkie-talkie equipment (b)

the microphone circuit are considerably magnified and capable of giving adequate frequency swing when the microphone is coupled to the oscillator circuit by the mutual inductance between L_2 and L_3.

REACTANCE-VALVE FREQUENCY MODULATORS

General

Most of the modulators used in f.m. transmitters make use of networks which include valves and have a predominantly reactive input impedance which can be varied by the modulating waveform. A simple example of such a circuit is that which is shown in Fig. 4.4. It contains a simple potential divider RC connected

across the input terminals of the network, the signal developed across C being applied between the grid and the cathode of a valve. For the sake of simplicity the h.t. and g.b. feeds for the valve are omitted. If the reactance of C is small compared with R at the frequency of an applied signal, the current in the potential divider is in phase with the signal. The voltage across C, i.e. the input signal to the valve, lags the current in the divider and hence the applied signal by 90°. The anode current of a valve is in phase with its grid-cathode signal and thus the anode current in Fig. 4.4 also lags the applied signal by 90°. A circuit which takes a lagging current is

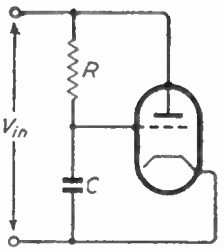


Fig. 4.4. Basic circuit of reactance modulator

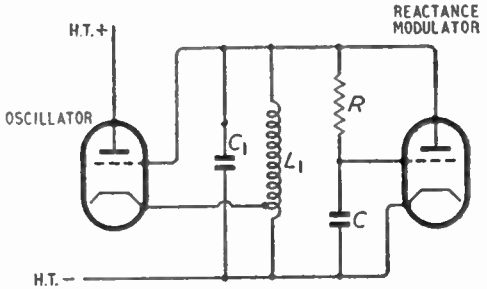


Fig. 4.5. A reactance modulator connected across an oscillating LC circuit

effectively inductive and thus the valve together with the RC circuit behaves as an inductive load on the signal source.

Moreover the current in the valve can be varied by alteration of its mutual conductance: such an alteration has the same effect on the current as a change of inductance. The circuit thus behaves as a variable inductance and can be used for frequency modulation by connecting the input terminals across the LC circuit of an oscillator (as shown in Fig. 4.5) and arranging for the modulating signal to control the mutual conductance of the valve. No source of h.t. or g.b. for the modulator valve is included in Fig. 4·5.

Theory of the Reactance-valve Modulator

It is essential that the input signal to the valve should be as nearly as possible in quadrature with the voltage across the circuit L_1C_1. If this is not arranged, the valve anode current contains a component in phase with the voltage across the LC circuit. This implies that the impedance presented by the valve anode circuit to the LC circuit has a resistive component. Such a resistance damps the tuned circuit, reducing its effective Q value, but—and

this is even more undesirable—the damping varies with the modulating signal and thus the voltage across the LC circuit is amplitude-modulated as well as frequency-modulated. To minimise this effect the reactance of C must be small compared with R at the operating frequency. The network RC can also damp the tuned circuit L_1C_1 in its own right, but fortunately the damping due to this source is also a minimum when the reactance of C is small compared with R.

The effective reactance of the valve anode circuit can be calculated quite simply in the following way. Let the alternating voltage across L_1C_1 be V_{in}. Then the current I in the potential divider is given by

$$I = \frac{V_{in}}{R + 1/j\omega C}$$

But $1/\omega C$ is small compared with R.

$$\therefore\ I = \frac{V_{in}}{R}$$

The voltage V_c developed across C by this current is equal to $I/j\omega C$

$$\therefore\ V_c = \frac{V_{in}}{j\omega CR}$$

The anode current of the valve is equal to $g_m V_c$, i.e.

$$I_a = \frac{g_m V_{in}}{j\omega CR}$$

The impedance presented by the anode circuit of the valve is given by V_{in}/I_a which, from the above expression, is given by

$$\frac{V_{in}}{I_a} = \frac{j\omega CR}{g_m}$$

This is equivalent to the reactance of an inductor L_e where

$$j\omega L_e = \frac{j\omega CR}{g_m},$$

giving

$$L_e = \frac{CR}{g_m}$$

The valve anode circuit behaves as an inductance proportional to

the time constant RC and inversely proportional to the mutual conductance of the valve.

As an alternative approach to the design of a reactance valve modulator, it is possible to arrange for the reactance of C to be large compared with R at the frequency of operation, the valve being connected across R. This gives the valve an input which leads the voltage across the tuned circuit by 90°, and the anode

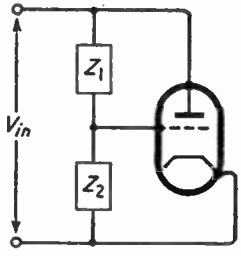


Fig. 4.6. General circuit for a reactance modulator

circuit behaves as a capacitance. By reasoning similar to that employed above we can show that the effective capacitance C_e of the valve anode circuit is given by

$$C_e = g_m RC$$

It is also possible to design reactance modulators in which the potential divider contains an inductor in series with a resistor. In fact we can represent the circuit of a reactance modulator in general as shown in Fig. 4.6 in which the two arms of the potential divider are labelled $\mathcal{Z}_1$ and $\mathcal{Z}_2$. Details of the four circuits chiefly employed in practice are given in the following table:

$\mathcal{Z}_1$	$\mathcal{Z}_2$	*Condition for satisfactory operation*	*Effective L or C of valve anode circuit*
R	C	$R \gg 1/j\omega C$	$L_e = RC/g_m$
C	R	$1/\omega C \gg R$	$C_e = g_m RC$
R	L	$R \gg \omega L$	$C_e = g_m L/R$
L	R	$\omega L \gg R$	$L_e = L/g_m R$

Design of a Single-valve Reactance Valve Modulator

In a numerical example given above it was shown that an oscillator operating at 10 Mc/s and tuned by a 50-pF capacitor requires a

capacitance change of $\pm 0{\cdot}5$ pF to give a frequency swing of ± 50 kc/s. We shall continue this numerical example to illustrate one way in which a reactance valve modulator can be designed.

We shall choose the circuit listed second in the above table, because it requires a capacitive potential divider (which is easier to construct than an inductive one) and because it gives an effective capacitance directly proportional to the mutual conductance of the valve. Thus, if we arrange to bias the valve to give a mutual conductance of say 1 mA/V in the absence of modulation, we can arrange for the modulating signal to swing the g_m between 1·5 mA/V and 0·5 mA/V on peak signals. The peak swing of mutual conductance Δg_m is thus 0·5 mA/V, and this produces a swing of effective capacitance given by

$$\begin{aligned} \Delta C_e &= \Delta g_m RC \\ &= 0{\cdot}5 \times 10^{-3} \times RC \end{aligned}$$

We know that the capacitance swing required is $\pm 0{\cdot}5$ pF. This determines the time constant thus

$$0{\cdot}5 \times 10^{-12} = 0{\cdot}5 \times 10^{-3} RC$$

$$RC = \frac{0{\cdot}5 \times 10^{-12}}{0{\cdot}5 \times 10^{-3}}$$

$$= 10^{-9}$$

We must now see whether this value of RC satisfies the requirement that the reactance of C is large compared with R at the operating frequency.

The ratio of reactance to resistance is given by

$$\frac{X}{R} = \frac{1}{\omega RC}$$

and as the frequency is 10 Mc/s we have

$$\frac{X}{R} = \frac{1}{6{\cdot}284 \times 10 \times 10^6 \times 10^{-9}}$$

$$= 17 \text{ approximately}$$

This value exceeds 10, and is satisfactory. Had it been less than 10 it would be necessary to recalculate for a smaller value of ΔC_e. Finally we have to determine values for R and C. We do not

wish too much capacitance to be thrown across the tuned circuit, and $C = 10$ pF is a reasonable value to accept. R is then given by

$$R = \frac{\text{time constant}}{C}$$

$$= \frac{10^{-9}}{10 \times 10^{-12}}$$

$$= 100 \text{ ohms}$$

Practical Circuit for Single-valve Reactance Modulator

In the calculation just performed it was necessary to change the mutual conductance of the modulator valve to give the required capacitance changes. This can be achieved by variation of grid

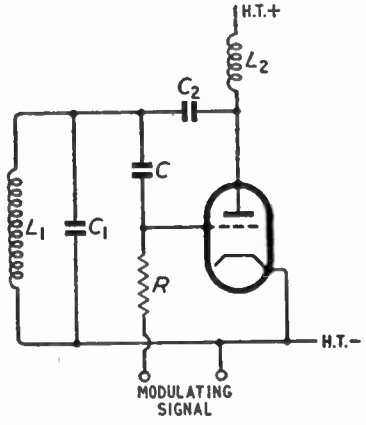


Fig. 4.7. Practical circuit for a single-valve reactance modulator

bias, and a circuit for a frequency modulator of this type is given in Fig. 4.7. This circuit includes a number of practical features not so far mentioned. For example, it is necessary to apply h.t. to the modulator valve, and the blocking capacitor C_2 is included to prevent a short-circuit of the h.t. supply by the inductor L_1. The anode cannot be connected directly to h.t. positive as this would place a low impedance path across L_1C_1. Thus an inductor L_2 is included in the anode circuit: this is effectively in parallel with L_1 and must be large compared with L_1 to avoid undue reduction in the effective inductance in the oscillator tuned circuit.

To achieve modulation in a circuit of this type the relationship between mutual conductance and grid bias must be linear: for

example, if the mutual conductance is 1 mA/V at a grid bias of, say, −2 volts and 1·5 mA/V at −1 volt, it must be 0·5 mA/V at −3 volts. A modulating signal of 1 volt peak then swings the mutual conductance between the limits of 0·5 mA/V and 1·5 mA/V. As valves are biased back towards anode-current cut-off the mutual conductance does fall, and for certain valves, notably triodes, the relationship between mutual conductance and grid bias is approximately linear over a limited range of grid voltage, giving acceptable linearity of modulation.

Suppressor-grid Modulation

An alternative method of changing the mutual conductance of a reactance modulator which uses a pentode is to apply the modulating voltage to the suppressor grid, as indicated in Fig. 4.8. The effect of applying a negative voltage to the suppressor grid is to decrease anode current and increase screen-grid current, the total current leaving the cathode remaining substantially unaffected. The anode-current suppressor voltage characteristic has a shape similar to that of the anode current-grid voltage characteristic of a

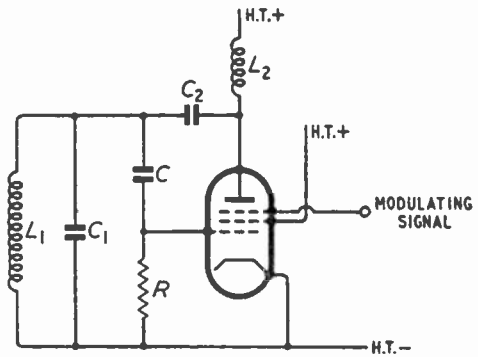


Fig. 4.8. A reactance modulator employing a pentode

triode, and if the suppressor grid is biased near the centre of the linear portion of this characteristic the effective mutual conductance of the pentode is roughly proportional to the suppressor-grid voltage.

The performance of simple reactance modulators of this type represented by Figs. 4.7 and 4.8 is, however, not good enough for a high-quality broadcast transmitter, largely because of the difficulty of obtaining the requisite frequency stability. Even in the absence of modulating signals, changes in h.t. voltage, l.t. voltage, or ageing of the valve result in variations of effective anode capacitance and

hence in oscillator frequency. Broadcast transmitters must adhere to their allotted carrier frequency within very close tolerance, and the frequency modulator employed must not be subject to such wide changes of frequency as occur in single-valve circuits.

Single-valve Reactance Modulators with Inductive Anode Circuit

A reactance valve modulator is connected in parallel with the LC circuit, the frequency of which it is required to control. The performance is easy to calculate if the modulator behaves as a capacitance (as in the modulators listed second and third in the table on page 58) because the capacitance is simply added to that of the LC circuit. If, however, the modulator is inductive (as in the modulators listed first and fourth), the relationship is a little more complicated because the total inductance in the tuned circuit is not the sum of the individual inductances but is given by $LL_e/(L + L_e)$. The resonance frequency is hence given by

$$f = \frac{1}{2\pi\sqrt{\left(\frac{LL_e}{L + L_e}\right)}}$$

Suppose the frequency changes to $f + \Delta f$ when the effective inductance changes to $L_e + \Delta L_e$ as a result of a change of mutual conductance.

$$\therefore f + \Delta f = \frac{1}{2\pi\sqrt{\left[\frac{L(L_e + \Delta L_e)}{L + L_e + \Delta L_e}\right]}}$$

Dividing the two equations we have

$$\frac{f + \Delta f}{f} = \sqrt{\left(\frac{LL_e}{L + L_e}\right)}\sqrt{\left[\frac{L + L_e + \Delta L_e}{L(L_e + \Delta L_e)}\right]}$$

$$= \sqrt{\left(\frac{L_e}{L_e + \Delta L_e}\right)} \text{ approximately}$$

$$= \sqrt{\left(\frac{L_e - \Delta L_e}{L_e}\right)}$$

$$= 1 - \frac{\Delta L_e}{2L_e}$$

$$\therefore \frac{\Delta f}{f} = -\frac{\Delta L_e}{2L_e}$$

The result is thus similar to that for capacitance control, and if ΔL_e is made proportional to g_m, linear modulation can be achieved.

Push-pull Reactance Modulators

One method of improving the frequency stability of a reactance modulator is to connect two valves in parallel across the LC circuit to be controlled, one valve behaving as a capacitance and the other as an inductance. The modulating signals are applied to the two valves in push-pull and the design is such that an increase in capacitance in one valve is accompanied by an increase in inductance in the other, the effect on the frequency shift being additive. The circuit of a push-pull reactance modulator of this type using suppressor-grid injection is given in Fig. 4.9. One valve has an

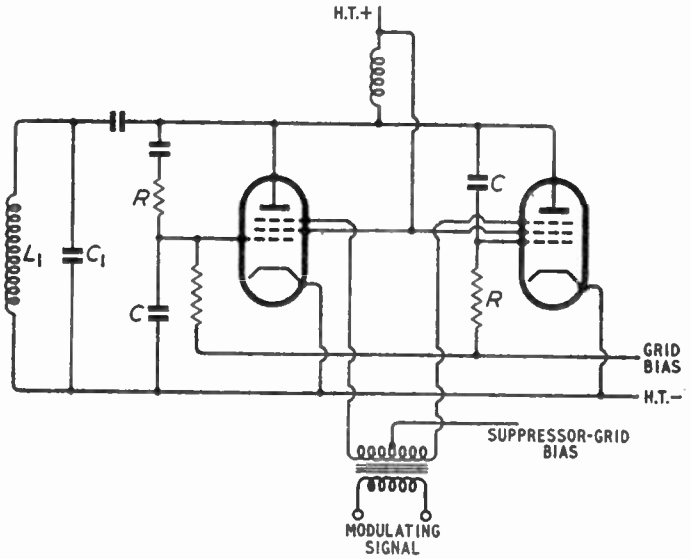


Fig. 4.9. Basic circuit for a push-pull reactance modulator using pentodes

RC potential divider of the type giving a capacitive anode circuit and the other a divider giving an inductive anode circuit.

Frequency drift due to variations in h.t. or l.t. voltage is much less in this than in single-valve modulators because such variations constitute signals applied in the same sense to the two valves and the circuit is sensitive only to potentials applied in push-pull. The

capacitance in the potential divider for one valve is often made variable and is adjusted empirically to give best stability. As the oscillator valve feeding the LC circuit to be controlled is itself likely to give frequency variations with variations in supply voltages, the capacitance may alternatively be set to give the best overall

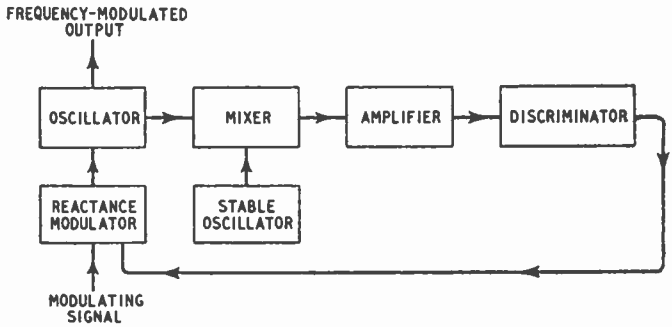


Fig. 4.10. Block diagram illustrating one method of improving the frequency stability of a reactance-modulated oscillator

stability from the combination of reactance modulator and oscillator. A second advantage of the push-pull circuit is that if the grid voltages are not precisely in quadrature with the anode voltages, the resulting variations in damping tend to cancel, thus minimising amplitude modulation of the oscillator.

A push-pull reactance modulator has low frequency drift, but better frequency stability is required in an f.m. broadcast transmitter and a number of circuits giving the required performance have been developed. What is wanted, of course, is frequency stability of the order of that obtained in an a.m. transmitter where it is customary to use an oscillator controlled by a thermostatically maintained quartz crystal. There are two possible ways in which such stability may be obtained.

One is to compare the centre frequency of the output from the frequency-modulator with the output of a quartz-controlled oscillator and to use a correcting circuit which automatically adjusts the centre frequency so as to minimise this difference. The other method is to attempt to frequency modulate a quartz-controlled oscillator directly.

In the first method the frequency-modulated oscillator may have a nominal centre frequency of, say, 5 Mc/s, and the stable oscillator

of, say, 7 Mc/s. If samples of both signals are applied to a mixer stage as indicated in Fig. 4.10, an output is obtained at the difference frequency, 2 Mc/s. This can be applied to an amplifier driving a discriminator tuned to 2 Mc/s, the output of which indicates the magnitude and direction of any wandering of the centre frequency of the frequency-modulated oscillator. If the output of the discriminator is applied to the reactance modulator as an error feedback signal, the variations of centre frequency can be reduced by a factor of 5 : 1 or more, giving stability good enough for a broadcast f.m. transmitter.

FMQ SYSTEM OF FREQUENCY MODULATION

Quartz crystals behave electrically as series LCR circuits with high inductance, low capacitance, and extremely low resistance. At or near the resonance frequency they present an impedance so low that conventional reactance modulators connected across the crystal have negligible effect on the resonance frequency. Conventional

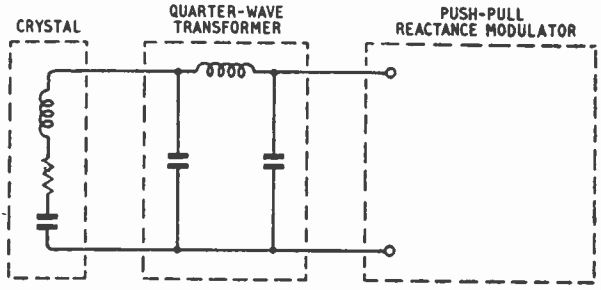


Fig. 4.11. Essential features of the FMQ system of frequency modulation

push-pull reactance modulators are effective only when connected across high-impedance circuits such as parallel LC circuits at or near their resonance frequency.

The difficulty of frequency-modulating a crystal can be overcome and high frequency stability achieved by inserting an impedance-transforming device between the crystal and the reactance modulator, the circuit being designed to present the modulator with a suitably high impedance. A π-section network is suitable provided the L and C values are correctly chosen: the network is then termed a quarter-wave transformer. The essential features of the resulting circuit are illustrated in Fig. 4.11: it is known as the FMQ

(frequency-modulated quartz) circuit, and is used in a large number of broadcast f.m. transmitters.

ARMSTRONG'S FREQUENCY MODULATOR

Another and quite different method of frequency-modulating a quartz crystal, due to Armstrong, is used in the U.S.A. The essential features of the method are as follows. A quartz-crystal-controlled oscillator is first amplitude-modulated by the modulating signal in a balanced circuit which removes the carrier component, leaving only the sidebands as output. This output is phase-shifted by 90° and mixed with the original carrier to give a phase-modulated wave. This may be shown by vector diagrams. Fig. 4.12(a) represents a carrier wave v_c together with two sideband vectors v_1 and v_2. We shall assume 100 per cent amplitude modulation for which the amplitude of each sideband is one half that of the carrier. The sideband vectors have angular frequencies of $(\omega_c + \omega_p)$ and

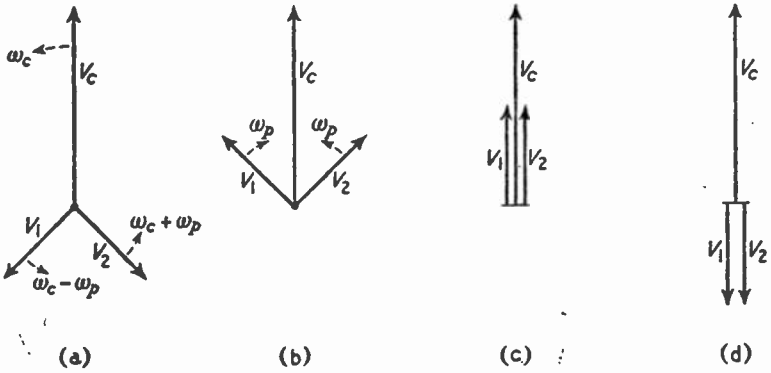


Fig. 4.12. Vector diagrams for a 100 per cent amplitude-modulated wave

$(\omega_c - \omega_p)$, where ω_p is the modulating angular frequency and ω_c is the carrier angular frequency.

If the carrier vector is assumed stationary, one sideband vector rotates at an angular frequency of $+\omega_p$ and the other at $-\omega_p$, i.e. they rotate in opposite directions as indicated in Fig. 4.12(b) in which the sideband vectors have rotated one quarter of a cycle relative to their position at (a). When the sideband vectors are in line as at Fig. 4.12(c) they combine to produce a resultant twice the length of the carrier vector. When they are in line as at Fig. 4.12(d) they neutralise the carrier vector to produce zero resultant.

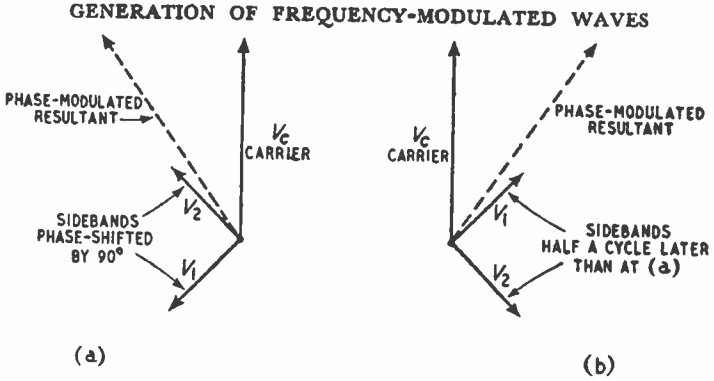


Fig. 4.13. Vector diagrams illustrating the action of Armstrong's phase modulator

The alternation in carrier amplitude between twice its unmodulated amplitude and zero confirms the initial assumption that the carrier is 100 per cent amplitude modulated.

Now let the sideband vectors be phase advanced by 90° relative to the carrier vector. This gives Fig. 4.13(a) in which the sideband vectors are 90° advanced relative to their positions in Fig. 4.12(b): Fig. 4.13(b) shows the vector diagram half a cycle of the modulating frequency later. By combining the two sidebands with the carrier, the resultant vector shown dotted in Fig. 4.13(a) and (b) is produced. These diagrams show that the phase change in the modulating signal produces a phase change in the resultant vector, and this change does not depend on the frequency of the modulating signal but only on its amplitude: thus this method of modulation produces a phase-modulated output. If a frequency-modulated output is required, the modulating signal must be integrated (i.e. passed through a network giving 6 dB lift per octave reduction in frequency) before application to the modulator.

DETAILS OF F.M. TRANSMITTERS

We have discussed in some detail some of the modulator circuits used in f.m. transmitters and some of the methods used to ensure the required linearity of modulation and frequency stability. One of the advantages of an f.m. over an a.m. transmitter is that the modulation process can be carried out at low power: in fact, the valves used in frequency modulators are seldom much larger than those used in receivers. Subsequent power amplifying stages, operating in Class C, are used to raise the carrier power to the level

required for radiation. The modulator usually operates at a low frequency which is subsequently multiplied in later stages to the carrier frequency. For example, if a carrier frequency of 96 Mc/s and a deviation of ± 75 kc/s are required, the modulator may operate at 8 Mc/s with a deviation of $\pm 6{\cdot}25$ kc/s, multiplication by 12 giving the required final values of carrier and deviation.

Basically, a frequency multiplier is a class-C amplifier, the anode circuit of which is tuned to a harmonic of the input signal. Such

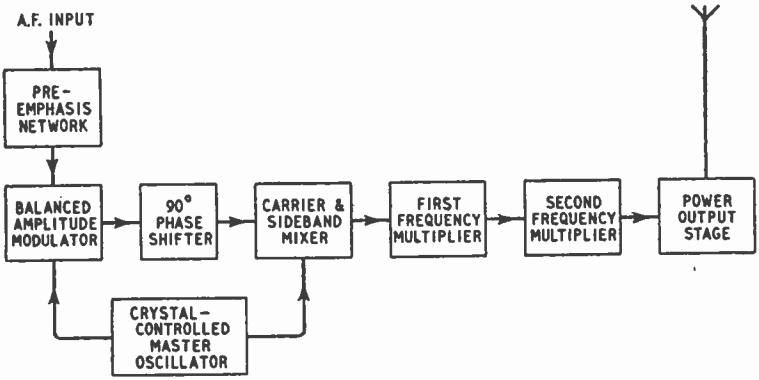


Fig. 4.14. Block diagram for an f.m. transmitter using Armstrong's phase modulator

stages have a grid bias greater than cut-off, and the anode current consists of a series of pulses at the frequency of the input signal. The pulses have a rich harmonic content, and the anode circuit may be tuned to any desired harmonic. However, an anode circuit tuned to, say, the tenth harmonic is required not to respond significantly to the ninth and eleventh harmonics, and this requires a high degree of selectivity which may be difficult to achieve. Multiplication factors are thus kept low, and the factor of 12 required in the chosen example would most probably be achieved in two stages of multiplication, one giving a factor of 3 and the other of 4. Usually the frequency multipliers follow the modulator directly, and one possible sequence of stages in a high-power f.m. transmitter follows the pattern indicated in the block schematic diagram of Fig. 4.14.

The performance of modern transmitters leaves little to be desired in respect of linearity and frequency response. The harmonic distortion does not appreciably exceed 1 per cent, and the response is usually flat within ± 1 dB from 30 c/s to 15,000 c/s.

5

DETECTION OF FREQUENCY-MODULATED WAVES

Introduction

THIS chapter describes circuits for deriving the modulation wave-form from a frequency-modulated wave. These circuits are termed detectors or discriminators, and one such circuit is the most important stage in an f.m. receiver.

It is difficult to design an f.m. detector to operate at the very high carrier frequencies (e.g. 90 Mc/s) used for f.m. transmission, and it is also difficult to obtain the gain necessary in a receiver at such frequencies. It is customary therefore to use the superheterodyne principle in reception with an intermediate frequency of the

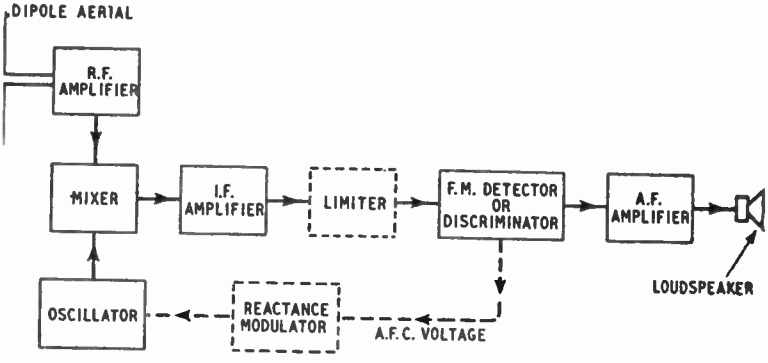


Fig. 5.1. Block diagram for a superheterodyne f.m. receiver

order of 10 Mc/s, detection and most of the amplification being carried out at this frequency.

A block schematic diagram for such a receiver is shown in Fig. 5.1. This shows a reactance valve in dotted lines: this can be omitted if adequate oscillator-frequency stability can be obtained without it. Also shown in dotted lines is a limiter stage before the detector. This may also be omitted if the detector is one of the self-limiting types, such as the Ratio Detector, described below.

SLOPE DETECTOR

The purpose of the f.m. detector is to produce a voltage substantially proportional to the instantaneous frequency of the input modulated wave. The basic need is for a device with a reasonably linear voltage-frequency characteristic. An inductor fed from a high-impedance source of f.m. signals can be used, but the output voltage generated across an inductor is small unless a large inductor is used, and this necessitates low values of intermediate frequency which introduce a number of difficult problems. Moreover, to obtain

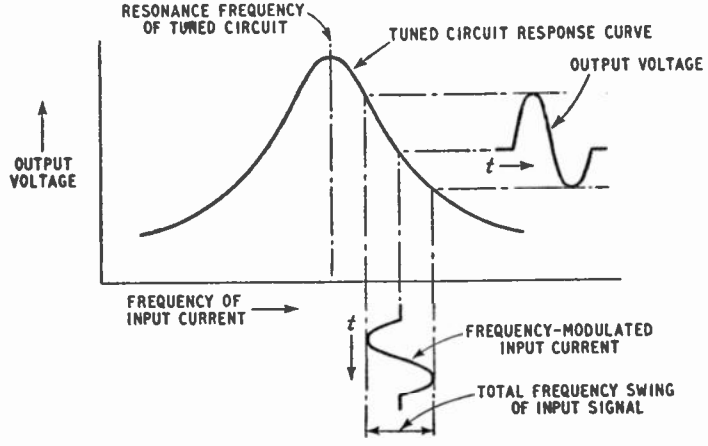


Fig. 5.2. Operation of slope detector

adequate output the reactance must change more rapidly with frequency than the reactance of a simple inductor.

Such a variation can be obtained from a combination of inductance and capacitance such as a parallel LC combination, the net reactance of which changes very rapidly as frequency approaches resonance. With an LC circuit of high Q value the slope of the reactance-frequency curve is high enough, even at 90 Mc/s, to give worthwhile output. Detectors utilising the slope of the skirts of a resonance curve are termed slope detectors and were at one time used in simple f.m. receivers.

The operation of a slope detector is illustrated in Fig. 5.2. In this example the LC circuit resonates below the centre frequency of the f.m. signal and frequency swings produce changes in output voltage: thus the output is a carrier wave simultaneously frequency- and amplitude-modulated. This output is applied to an a.m.

detector: such a detector ignores frequency fluctuations and its output is proportional to the amplitude-modulated content.

The tuned circuit has, of course, two regions which can be used for slope detection, and if there are no selective circuits ahead of the detector each f.m. signal can be received at two settings of the tuning control. If, however, the detector follows a selective amplifier such as an i.f. amplifier, the design can be such that signals applied to unwanted portions of the characteristic are greatly attenuated: in this way repeat tuning points are eliminated.

Perhaps the most serious disadvantage of the slope detector is the harmonic distortion resulting from the inevitable curvature of the tuned-circuit characteristic.

ROUND-TRAVIS DETECTOR

One of the methods employed in amplifiers to reduce distortion caused by the curvature of valve characteristics is to employ the push-pull principle: this eliminates even-order harmonics. This principle can be applied to the slope detector by using two LC circuits, one resonant above and the other below the centre frequency of the signal to be detected. A detector having these principles is termed a Round-Travis detector.

A typical circuit is given in Fig. 5.3. The input is assumed to be

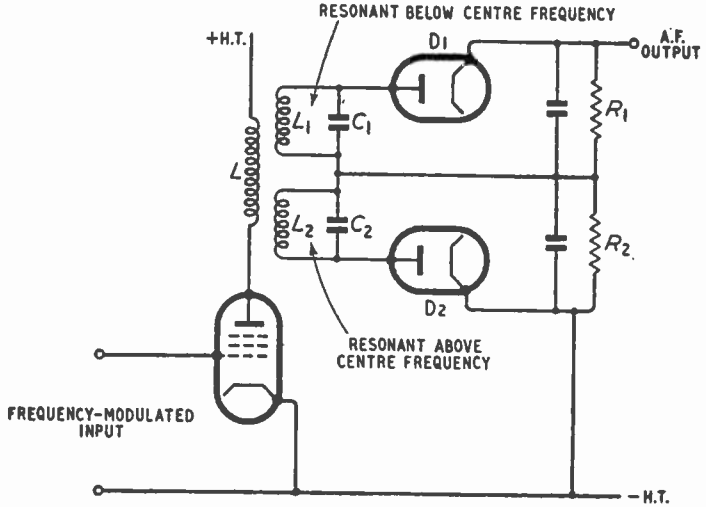


Fig. 5.3. Circuit for a Round-Travis detector

in the form of a current flowing in the primary winding L of the transformer. This is coupled to the tuned circuit L_1C_1 resonant below the centre frequency, the output of which is rectified by the diode D1 to produce an output across the load R_1. The primary winding is also coupled to the circuit L_2C_2 resonant above the centre frequency, the output of which is rectified by D2 to produce an output across R_2. The output loads R_1 and R_2 are connected in series, and the net output across $(R_1 + R_2)$ constitutes the detector output.

The two circuits L_1C_1 and L_2C_2 are similar in construction and their resonance frequencies are equally displaced from the centre frequency: thus, if the input signal is at the centre frequency the signals applied to the two diodes are equal, the voltages across R_1 and R_2 are equal, and the net output is zero. If the frequency of the input signal is decreased, it approaches the resonance frequency of L_1C_1 which delivers a larger signal to D1 and gives a larger voltage across R_1. The decrease in signal frequency causes the voltage across L_2C_2 to fall, and the voltage across R_2 also falls. Thus, there is a net voltage across $(R_1 + R_2)$.

Similarly, when the signal frequency swings above its centre value, the voltage across L_2C_2 increases and the voltage across L_1C_1 falls. The voltage across R_2 now exceeds that across R_1 and the net output voltage is opposite in sign to that given by a decrease in signal frequency. Thus, the detector output signal indicates by its polarity whether the frequency displacement is upward or downward and indicates by its magnitude the extent of the frequency displacement.

The operation of the Round-Travis detector is illustrated in Fig. 5.4 which corresponds with Fig. 5.2. L_1C_1 is resonant at a frequency f_1 and L_2C_2 at a frequency f_2, the centre frequency being at $(f_1 + f_2)/2$. One resonance curve is shown inverted with respect to the other to indicate that the two diode outputs are added in phase opposition.

The effective characteristic of the two LC circuits acting together is shown by the dotted line and this is a much better approximation to the ideal linear response than the curve of a single LC circuit shown in Fig. 5.2.

The dotted characteristic with its linear centre portion has a shape similar to that of an elongated letter S lying horizontally. This type of input frequency–output voltage characteristic is typical of an f.m. detector: similar responses are obtained from Foster-Seeley and Ratio-Detector circuits.

To keep distortion low, the centre part of the detector output-frequency curve should be linear. The linearity of this part of the curve depends on the working Q values of the tuned circuits and the difference between their resonance frequencies. The greater the Q values the smaller must be the frequency difference for good

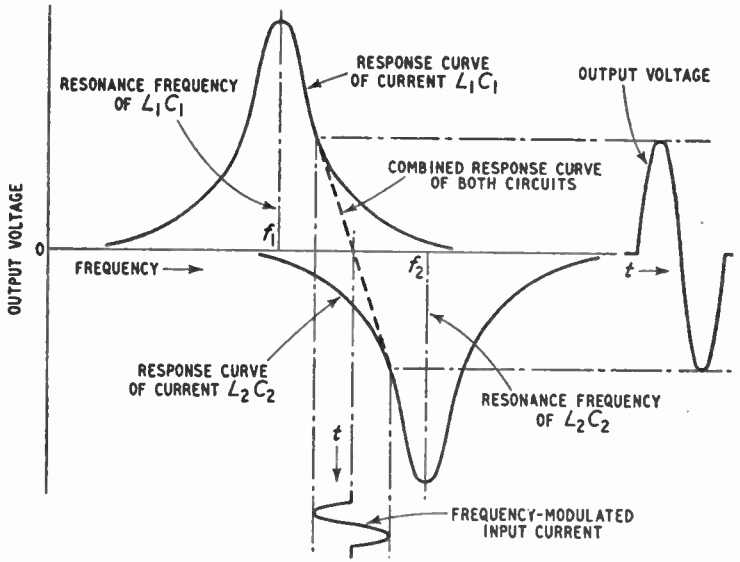


Fig. 5.4. Operation of a Round-Travis detector

linearity. Results are usually found to be acceptable if the ratio of frequency difference to centre frequency is between $1/Q$ and $1{\cdot}5/Q$.

The performance of the Round-Travis circuit is greatly superior to that of the slope detector. Nevertheless, it has not been used in f.m. receiver design to any extent. This is because the adjustment of the circuit is more difficult than that of other circuits described below.

FOSTER-SEELEY DISCRIMINATOR

This form of f.m. detector is in many respects similar to the Round-Travis circuit, but it is simpler to adjust because both LC circuits are aligned at the centre frequency of the signal. One form of

Foster-Seeley circuit is illustrated in Fig. 5.5 which shows that it has two diodes D1 and D2 with load resistors R_1 and R_2 which are connected in series as in the Round-Travis circuit. The method of producing the signals for the diode anodes is, however, quite different. The diodes are fed from a secondary tuned circuit L_2C_2, coupled to a primary tuned circuit L_1C_1.

An essential feature of the Foster-Seeley circuit is that a fraction of the voltage generated across the primary circuit is applied to the centre point of the secondary circuit. In Fig. 5.5 the connection is between the centre point of L_2 and a tapping point on L_1, but the secondary connection may be to the junction of two equal capacitors across L_2 (they could together constitute the tuning capacitance) and the primary connection may be to an inductor closely coupled to L_1 or to a capacitive potential divider across L_1 (formed by capacitors which may also provide the tuning capacitance).

The Foster-Seeley circuit relies for its operation on the phase relationships between the signals in a pair of coupled circuits and is best explained by vector diagrams.

Suppose a current at the resonance frequency of L_1C_1 and L_2C_2 is applied to L_1C_1 from a high-resistance source which does not appreciably damp the circuit. A parallel LC circuit is resistive at resonance and the voltage generated across L_1C_1 is in phase with the

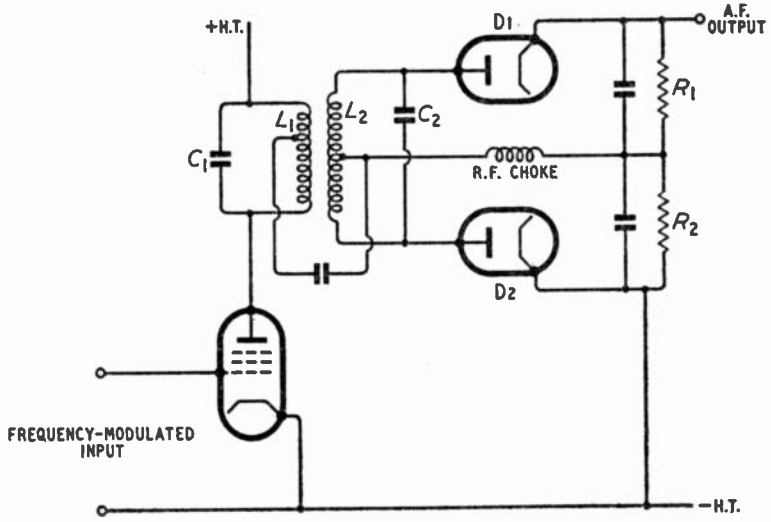


Fig. 5.5. Circuit for a Foster-Seeley discriminator

initial current as shown in Fig. 5.6. The current in L_1 lags the signal across L_1 by 90°, and the e.m.f. induced in L_2 by this current (being given by $j\omega Mi$) leads this current by 90° and is thus in phase with the initial current. L_2C_2 is resonant at the frequency of the applied signal and is hence resistive: thus the induced e.m.f.

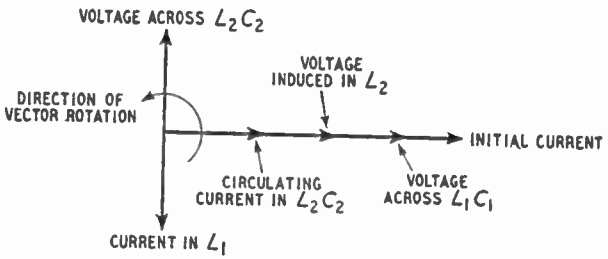


Fig. 5.6 (above). *Vector relationships in the circuit of Fig. 5.5, when the applied signal is at the resonance frequency of L_1C_1 and L_2C_2*

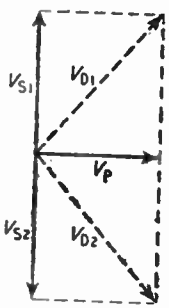


Fig. 5.7 (right). *Vector relationships in the circuit of Fig. 5.5 when the applied signal is at the resonance frequency of L_1C_1 and L_2C_2*

drives through the series circuit comprising L_2 and C_2 a circulating current which is in phase with the induced e.m.f. and hence with the initial current. This circulating current produces a voltage across C_2 (or L_2) which is in quadrature with the circulating current and hence with the initial current. Finally—and this is the principle on which the circuit depends for its action—the voltage across L_2C_2 is in quadrature with the voltage across L_1C_1.

The input to the diode D1 is made up of the voltage across the upper half of the secondary circuit L_2C_2 together with the voltage from the primary circuit. These are represented by the vectors V_{s1} and V_p in Fig. 5.7. Similarly, the input to the diode D2 is made up of the voltage V_{s2} across the lower half of the secondary circuit (which is in phase opposition to V_{s1}) and the same primary voltage V_p from the primary circuit. At the resonance frequency of L_1C_1 and L_2C_2, V_{s1} and V_{s2} are equal in magnitude and both are in

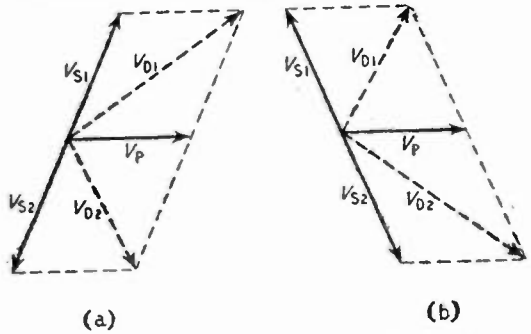


Fig. 5.8. Vector relationships in the circuit of Fig. 5.5, when the frequency of the input signal is (a) above and (b) below the resonance frequency of L_1C_1 and L_2C_2

quadrature with V_p. By adding V_{s1} and V_p vectorially we find that the input to D1 is represented in magnitude and direction by the vector V_{D1}. The input to D2 is made up of V_{s2} and V_p, the vector sum of which is represented by the vector V_{D2}.

Provided V_{s1} and V_{s2} are equal (implying an accurate centre tap of circuit L_2C_2) and provided V_{s1} and V_{s2} are in quadrature with V_p (implying that the input signal is at the resonance frequency of L_1C_1 and L_2C_2) V_{D1} and V_{D2} are equal, the voltages generated across R_1 and R_2 are equal, and there is no resultant voltage across $(R_1 + R_2)$.

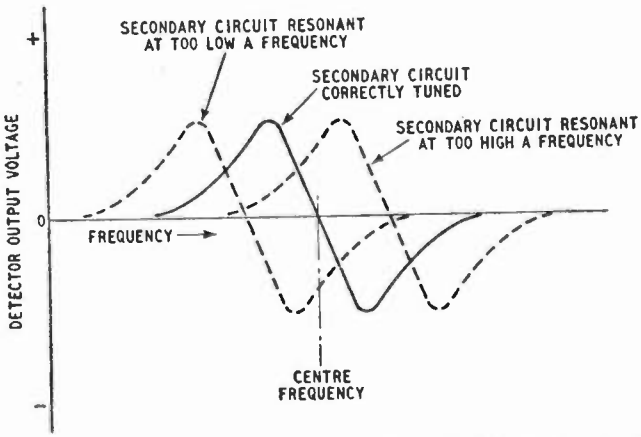


Fig. 5.9. Effect on f.m. detector response curve of mistuning the secondary LC circuit

If the frequency of the input signal is increased, the phase relationships change, V_{s1}, V_{s2} and V_p being positioned now as indicated in Fig. 5.8(a). The resultant vectors V_{D1} and V_{D2} are no longer of equal magnitude and the voltage across R_1 and R_2 are unequal, giving a net output.

If the frequency of the applied signal is below the resonance frequency of L_1C_1 and L_2C_2 the phase relationships are as indicated in Fig. 5.8(b). Again the resultant vectors are unequal and there is a resultant voltage across $(R_1 + R_2)$, this time of opposite polarity to that given by an input frequency above the resonance value.

If the input is frequency-modulated with a centre frequency equal to the resonance frequency of L_1C_1 and L_2C_2, the output voltage across $(R_1 + R_2)$ swings positively and negatively with a

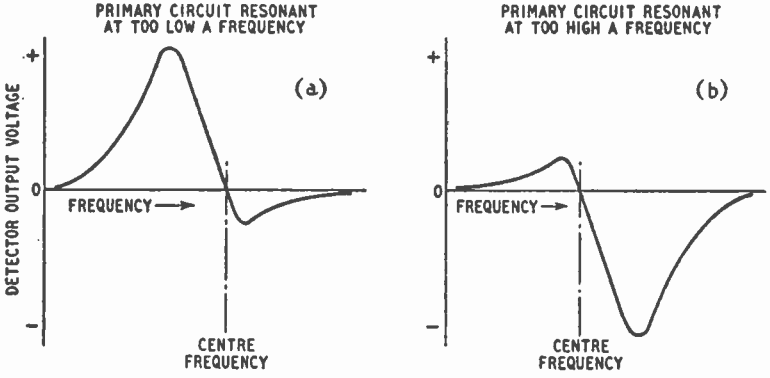


Fig. 5.10. Effect on f.m. detector response curve of mistuning the primary LC circuit

waveform which is a substantially faithful copy of the modulating signal.

The relationship between output voltage and frequency for a correctly adjusted Foster-Seeley discriminator has the elongated S shape already described for the Round-Travis circuit shown again in Fig. 5.9. If the resonance frequency of the secondary circuit L_2C_2 is mistuned, the discriminator characteristic moves bodily along the frequency axis without appreciable change of shape as shown in Fig. 5.9, the frequency of zero output being substantially the resonance frequency of the secondary circuit. If the resonance frequency of the primary circuit L_1C_1 is mistuned, the discriminator characteristic becomes asymmetrical as is indicated in Fig. 5.10.

To minimise distortion in the detector it is essential to obtain a linear relationship between output voltage and frequency at the centre of the characteristic.

The linearity depends primarily on the coupling between the primary and secondary windings L_1 and L_2 and on the ratio of the secondary voltage to the voltage which is derived from the primary circuit.

For a given ratio of secondary to primary voltage, the shape of the output voltage-frequency characteristic varies with primary to

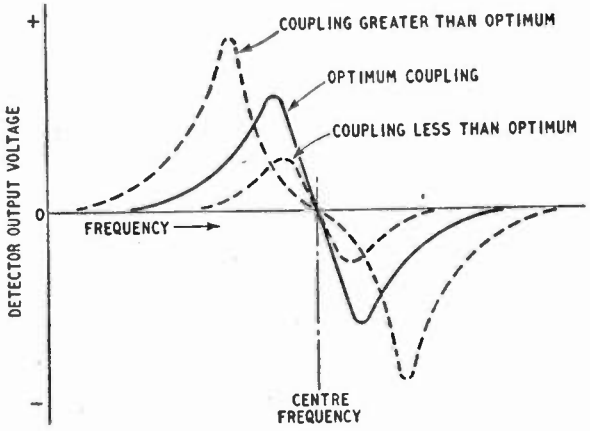


Fig. 5.11. Effect on f.m. detector response curve of altering the coupling between primary and secondary windings

secondary coupling as shown in Fig. 5.11. For small degrees of coupling the output increases as the coupling increases, but characteristic curvature becomes marked above a certain coupling. Linearity does not greatly vary at small values of coupling but deteriorates rapidly above this particular value. The optimum degree of coupling is thus the greatest which gives acceptable linearity.

NECESSITY FOR LIMITING

One of the advantages of frequency modulation over amplitude modulation is that it makes possible a great reduction in interference. To realise this advantage, however, an f.m. receiver must be relatively insensitive to amplitude-modulated signals. For good results an a.m. suppression ratio of at least 30 dB and preferably

40 dB is desirable. The a.m. suppression ratio, it will be recalled, is the ratio of the detector output to an a.m. signal, relative to that to an f.m. signal when the detector is fed with a signal simultaneously frequency-modulated and amplitude-modulated by different audible frequencies.

AMPLITUDE SUPPRESSION RATIO OF ROUND-TRAVIS AND FOSTER-SEELEY CIRCUITS

The Round-Travis and Foster-Seeley circuits give zero output at the centre frequency and are thus insensitive to amplitude modulation of a carrier at the centre frequency of the detector. If, however, the a.m. carrier is displaced from the centre frequency, the detector gives an output proportional to the modulation depth and to the frequency displacement. If the input signal is simultaneously frequency-modulated and amplitude-modulated then the detector responds to both components. The form of the output signal can be calculated in the following way.

Suppose the input signal is frequency-modulated only, the modulation frequency being $\omega_f/2\pi$. Then the detector output can be represented by $M \cos \omega_f t$, where M is the modulation index. If the input signal is amplitude-modulated as well, the detector output is given by

$$(1 + m \cos \omega_a t) M \cos \omega_f t,$$

where $\omega_a/2\pi$ is the amplitude-modulating frequency and m is the modulation factor.

Expanding this expression we have

$$(1 + m \cos \omega_a t) M \cos \omega_f t = M \cos \omega_f t + \frac{Mm}{2} \cos (\omega_f + \omega_a) t + \frac{Mm}{2} \cos (\omega_f - \omega_a) t$$

Of these three components the first represents the wanted f.m. signal and the other two are unwanted components at the sum and difference of the two modulating frequencies. This represents cross-modulation in the detector because the unwanted signals are not at the amplitude-modulating frequency. The unwanted components have an amplitude of $Mm/2$ and their power is thus proportional to

$$2 \times \frac{M^2 m^2}{4} = \frac{M^2 m^2}{2}$$

The power due to the f.m. signal is proportional to M^2 and the ratio

$$\frac{\text{power output due to amplitude modulation}}{\text{power output due to frequency modulation}} = \frac{M^2m^2}{2M^2}$$

$$= \frac{m^2}{2}$$

Thus the amplitude suppression ratio is given by

$$10 \log_{10} \frac{m^2}{2}$$

For measurement purposes m is usually made 40 per cent. Thus

$$\text{amplitude suppression ratio} = 10 \log_{10} \frac{m^2}{2}$$

$$= 10 \log_{10} \left(\frac{0{\cdot}4^2}{2}\right)$$

$$= 10 \log_{10} 0{\cdot}08$$

$$= -9 \text{ dB}$$

M does not enter into this ratio but is usually made 40 per cent also ($\pm$30 kc/s) because this represents average modulation. An a.m. suppression ratio of this value is not sufficient to give adequate protection against interference, and an f.m. receiver with a Round-Travis or Foster-Seeley circuit must have a limiter stage before the detector stage.

GRID LIMITER

The purpose of the limiter stage is to give an output with an amplitude independent of that of the input signal. The output thus contains the phase or frequency swings constituting the wanted signal but has constant amplitude and is thus free of noise or interference. The form of input-output characteristic required in a successful limiter stage is one which has a long straight horizontal section implying that output is independent of input. One circuit which gives a reasonable approximation to such a characteristic is that of the grid limiter, shown in its simplest form in Fig. 5.12. It resembles an i.f. amplifier but the valve is operated at a low anode and screen voltage (commonly around 40 volts) to keep the anode current low and to reduce the magnitude of the input signal required for limiting. The control grid and cathode are connected

as a grid-leak detector and across the grid resistor a steady voltage is generated approximately equal to the peak value of the input signal to the grid. This voltage biases the valve beyond the point of anode-current cut-off and pulses of anode current flow only on the peaks of each cycle of input signal as shown in Fig. 5.13.

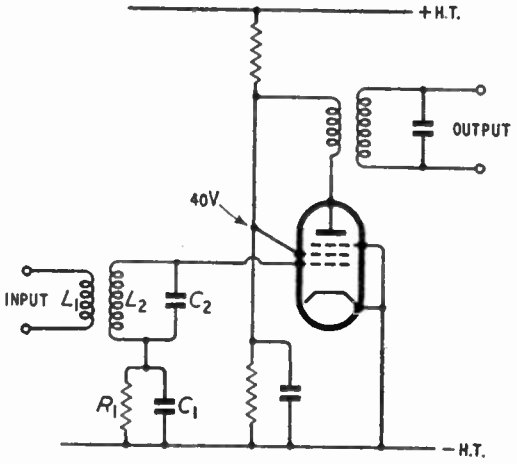


Fig. 5.12. Basic circuit of a grid limiter

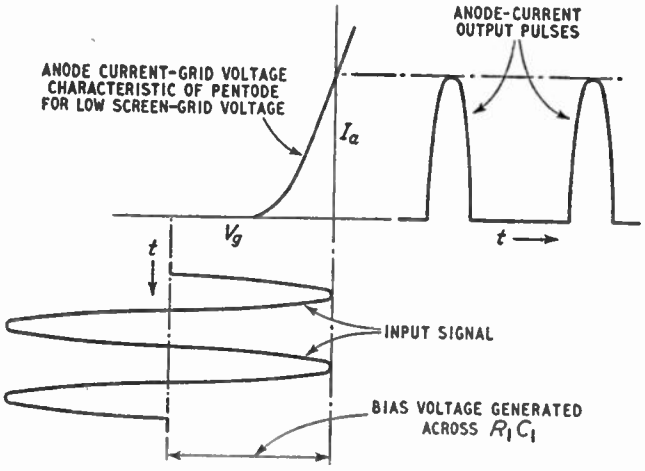


Fig. 5.13. Anode-current pulses in a grid limiter

The limiting action is illustrated in Fig. 5.14 which shows the waveform of the anode current pulses for three different amplitudes of input signal. The current pulses are of substantially the same amplitude for all three signals but the pulses become narrower, i.e. of shorter duration as the input amplitude increases. This has a significant effect on the limiter output voltage. The anode circuit contains an LC combination resonant at the input-signal frequency. As the current pulses become narrower the fundamental frequency component becomes smaller. Thus the output signal amplitude-input signal amplitude relationship has the form shown in Fig. 5.15.

For input amplitudes less than the grid base of the pentode there is no limiting and the curve shows a linear rise, indicating that for

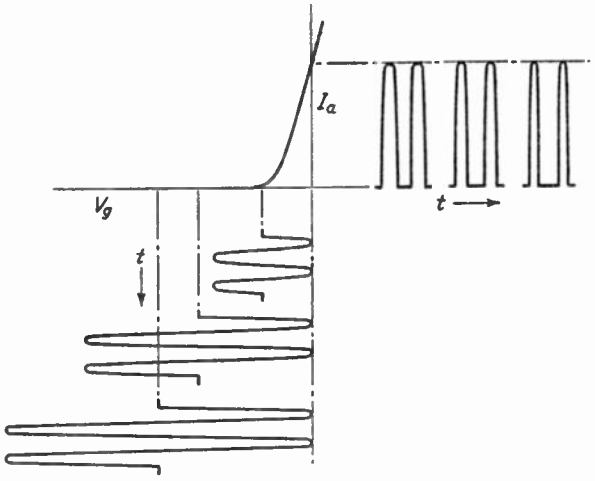


Fig. 5.14. Anode-current pulses for three different input-signal amplitudes

such small signals the output voltage is proportional to the input voltage. When the input amplitude equals the grid base, limiting begins and the curve levels off. Further increase in input amplitude causes the anode pulses to become narrow and the output signal falls slightly. This fall is undesirable because it implies lack of constancy in the output signal amplitude, but it can be offset by careful choice of anode and screen potentials.

For large input signals this form of limiter can give an amplitude modulation reduction ratio of up to 30 dB, that is to say variations in

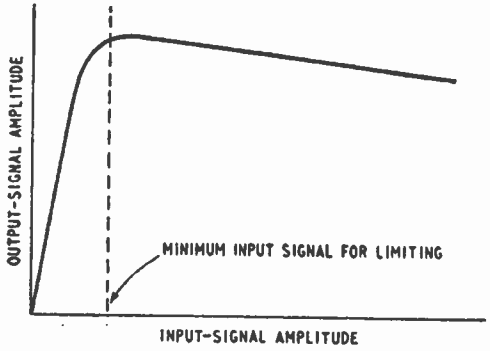


Fig. 5.15. Input-output characteristic for a grid limiter

output signal are only 1/30th those of the input signal. The circuit is extensively employed in f.m. receivers, usually in combination with a Foster-Seeley detector. The total a.m. suppression ratio of the limiter-detector combination exceeds 40 dB, representing a satisfactory protection against interfering signals. Such an arrangement can deal effectively with slow variations in input signal such as those caused by aircraft, and thus eliminates aircraft flutter. To deal with more rapid changes in input signal, such as those due to impulsive interference, the grid circuit time constant R_1C_1 must be made short.

There are, however, lower limits to both resistance and capacitance. If the resistance is made too small there is excessive damping of the input tuned circuit, causing low gain from the previous stage. This consideration sets a lower limit on the resistance R_1 of the order of 47,000 ohms. This reduces the Q value of a typical 10·7 Mc/s tuned circuit to approximately one half the undamped value. The capacitance C_1 must be large compared with the input capacitance of the pentode, otherwise there is an effective reduction in the valve input signal due to the potential divider formed by these two capacitances. A typical value for the valve input capacitance is 10 pF, and the grid capacitor cannot be much less than 50 pF. For these values of resistance and capacitance the grid circuit time constant is given by

$$\begin{aligned} R_1C_1 &= 47 \times 10^3 \times 50 \times 10^{-12} \\ &= 2{\cdot}5 \times 10^{-6} \text{ sec} \\ &= 2{\cdot}5 \ \mu\text{sec approximately} \end{aligned}$$

The chief disadvantage of this circuit is that limiting ceases for signals with an amplitude less than the grid base of the valve. This input is known as the threshold value and is commonly approximately 1 volt. Thus signals must exceed 1-volt amplitude and the amount by which they must exceed it depends on the downward amplitude modulation with which the limiter is required to deal. If it is considered desirable that downward amplitude swings of 90 per cent should be counteracted (and this may be necessary in areas where there is considerable multi-path propagation) then the input signal should be at least 10 volts in peak value. In other areas an input signal of only 3 volts peak value may give satisfactory reception, the circuit then being able to cope with downward amplitude swings of 67 per cent. Whether the input required is 3 volts or 10 volts it is obvious that high i.f. gain is required ahead of the grid limiter, and receivers incorporating such limiters must necessarily have more valves than those using the alternative types of limiter such as the ratio detector.

DYNAMIC LIMITER

Another type of limiter which can be used in f.m. receivers is the dynamic limiter. This circuit is not particularly important in its own right but will be described because similar principles are used in the ratio detector which is described afterwards. The ratio detector is widely used as a sound detector in American television receivers and in British sound receivers. Its performance is not quite up to the standard of the Foster-Seeley circuit but it can be designed to have a greater degree of inherent a.m. suppression. It is thus possible to dispense with a separate limiter in a receiver employing a ratio detector and this is a consideration of importance to radio receiver manufacturers who are faced with the problem of producing f.m. receivers economically. High-quality f.m. receivers, however, usually have a Foster-Seeley discriminator with one or two separate limiter stages.

One of the disadvantages of the grid limiter is that no limiting occurs for input signals with an amplitude less than a certain (threshold) value; the dynamic limiter has no such threshold. The basic circuit is illustrated in Fig. 5.16; it is similar to that of a diode detector. The circuit L_2C_2 is assumed resonant at the frequency of the current in the primary winding L_1.

When the circuit is used as an a.m. detector, the time constant R_1C_1 of the load circuit must be small enough for the voltage across R_1C_1 to change at an audible rate: the time constant must be

small when it is compared with the period of the highest audio frequency. On the other hand the time constant must not be too small, otherwise there is a loss of a.f. output. For maximum a.f. output the capacitor C_1 must be charged to the instantaneous value of the carrier voltage across L_2C_2 and this requires that R_1C_1 must be large compared with the period of the carrier. R_1C_1 has thus two requirements to satisfy and it is easily possible to find a value which satisfies both.

This gives the value of the product R_1C_1. The individual values of R_1 and C_1 can be determined by damping considerations. In a diode detector the damping across L_2C_2 is approximately equivalent to that of a resistance equal to $R_1/2$. When an i.f. amplifier is subjected to interfering signals of an impulsive nature, such signals

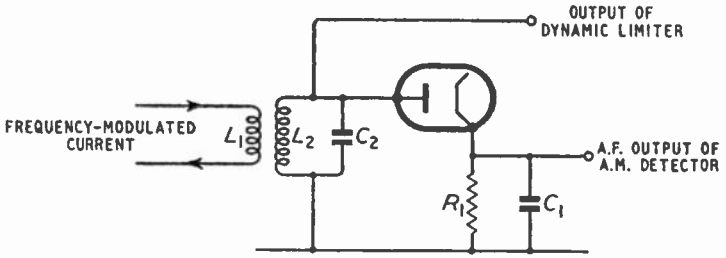


Fig. 5.16. Basic circuit for a diode a.m. detector or dynamic limiter

produce at the amplifier output damped wavetrains the shape of which is independent of the characteristics of the originating signals and depends only on the characteristics of the amplifier. When such wavetrains are applied to an a.m. detector, such as that shown in Fig. 5.16 the carrier envelope is faithfully reproduced in the voltage across C_1, showing that an a.m. receiver gives no protection against the effects of impulsive interference.

Now suppose the capacitance C_1 is increased to a value which makes the diode-load time constant large compared with the lowest audio frequency. This, of course, makes the circuit useless as an a.m. detector because the voltage across C_1 cannot now change quickly enough to reproduce audio frequencies. With such a value of C_1, however, the circuit can be used as a limiter in an f.m. receiver, the output being taken from L_2C_2 and not from R_1C_1.

Consider the behaviour of the circuit to an input current consisting of a constant-amplitude carrier at the resonant frequency of L_2C_2. The voltage across C_1 is still approximately equal to the

carrier amplitude across L_2C_2 and the damping imposed on the circuit L_2C_2 is still approximately equal to $R_1/2$. The only effect of the increase in the value of C_1 is that the voltage across C_1 takes longer to reach its final value after the carrier current is applied.

If the input to this circuit is subjected to impulsive interference, the current in L_1 contains damped wavetrains. Consider the behaviour of this circuit to a momentary increase in carrier amplitude. This tends to give an increased carrier voltage across L_2C_2. The voltage across C_1 can change only slowly, and any sudden increase in voltage across L_2C_2 causes the diode to take a larger current from the tuned circuit. In other words, on each positive half-cycle the diode takes more energy from the tuned circuit, which is therefore more heavily damped than before the carrier increase. This reduces the Q value of the circuit L_2C_2 and its dynamic resistance, with the result that the voltage across L_2C_2 does not increase in direct proportion to the increase of carrier input. In practice the increase in voltage may be only 1/6th of the increase in carrier input.

In this way the circuit is able to reduce the effects of a momentary increase in carrier amplitude and the reduction is obtained no matter how great the instantaneous value of carrier amplitude. If the increase in carrier amplitude were permanent, the voltage across C_1 and the voltage across L_2C_2 would rise exponentially to a new value proportional to the new carrier input, the rate of rise being determined by the time constant R_1C_1.

Now let us consider the behaviour of the circuit towards momentary decreases in carrier amplitude. When the input current falls, the voltage across L_2C_2 also falls. The voltage across R_1C_1 cannot fall quickly because of the large time constant, and the current taken by the diode is reduced. This reduces the damping of the circuit L_2C_2 and, if it was heavily damped previously, its Q value now rises, giving an increase in the dynamic resistance. Thus the voltage across L_2C_2 does not fall in direct proportion to the fall in carrier input, and in practice it is again possible to achieve a reduction in voltage of 1/6th the reduction in input current.

The limiting achieved on reductions of carrier input are only possible, of course, if the Q of the circuit can rise as a result of decreased damping. If the Q ever approaches its full (undamped) value, no further fall in carrier input can be counteracted. Thus there is a limit to the extent of downward-going amplitude modulation with which the circuit can deal. If the reduction in carrier amplitude is permanent, then the voltage across C_1 and across L_2C_2

fall exponentially to a value proportional to the reduced carrier input, the fall having a time constant of R_1C_1.

Since the circuit always adjusts itself so that the voltage across R_1C_1 is proportional to the carrier input, the circuit will limit on any value of input carrier: there is no threshold, as in the grid limiter. In practice, limiting becomes less effective as the carrier input is reduced, because of reduced diode efficiency.

Although the dynamic limiter is able to effect a considerable reduction in short-term carrier-amplitude changes, it cannot cope with slow variations in signal such as aircraft flutter. The capacitor C_1 charges and discharges rapidly enough to follow such variations in carrier, but limiting is achieved only when the voltage across C_1 remains constant. Aircraft flutter can be reduced by further increase of C_1 but the increase of the time constant beyond approximately 0·1 sec makes an f.m. receiver difficult to tune, for it is then possible to tune through signals without hearing them.

To obtain an estimate of the component values likely to be used in a dynamic limiter we may proceed as follows. At the intermediate frequency of 10·7 Mc/s commonly employed in f.m. receivers, resonant circuits are often tuned by 50 pF capacitors and the inductors have a Q of 70. The dynamic resistance of such a circuit is given by

$$R_d = \frac{Q}{\omega C}$$

$$= \frac{70}{6{\cdot}3 \times 10{\cdot}7 \times 10^6 \times 50 \times 10^{-12}} \text{ ohms}$$

$$= 30{,}000 \text{ ohms approximately}$$

If this tuned circuit is used to feed a dynamic limiter, for successful performance the dynamic resistance must be considerably reduced by damping due to the diode. Suppose the resistance is reduced to 1/10th, i.e. 3,000 ohms. The resistance necessary to do this is given by R in the equation

$$\frac{1}{R} + \frac{1}{30{,}000} = \frac{1}{3{,}000}$$

This gives

$$R = \frac{30{,}000 \times 3{,}000}{30{,}000 - 3{,}000} \text{ ohms}$$

$$= 3{,}300 \text{ ohms}$$

To give an effective damping resistance of this value, the diode load resistor R_1 must be double this, i.e. 6,600 ohms. If the time constant of the diode load circuit is 0·1 sec, the capacitor C_1 can be obtained from the relationship

$$R_1C_1 = 0{\cdot}1$$

$$C_1 = \frac{0{\cdot}1}{R_1}$$

$$= \frac{0{\cdot}1}{6{,}600} \text{ F}$$

$$= 13\ \mu\text{F approximately}$$

A.M. REDUCTION

A dynamic limiter with component values such as those just calculated should be capable of limiting even when the carrier input falls to 10 per cent of its initial value. It should also be capable of limiting no matter to what extent the carrier amplitude increases. It should therefore operate satisfactorily with amplitude modulation of the carrier up to 90 per cent depth. In a practical design such an input gives an output in which the amplitude modulation is reduced to ± 15 per cent. This is a 6 : 1 reduction, equivalent to an a.m. reduction of 15 dB. A.m. reduction should not be confused with a.m. suppression ratio which relates the output of a detector to an a.m. signal with its output to an f.m. signal when the input is a carrier simultaneously amplitude- and frequency-modulated.

RATIO DETECTOR

General

There are a number of f.m. detectors with a greater degree of inherent a.m. suppression than the Foster-Seeley circuit, and by far the most popular of these is the ratio detector, first described by Seeley and Avins in *R.C.A. Review* for June 1947. The basic circuit for one form of ratio detector is given in Fig. 5.17.

It contains two tuned circuits L_1C_1 and L_2C_2 both resonant at the centre frequency and coupled together as in a Foster-Seeley circuit. The two diodes are thus fed with signals which vary in frequency in precisely the same way as in a Foster-Seeley circuit. In the ratio detector, however, the two diodes are arranged in a "series-aiding" arrangement and feed a common load circuit

R_3C_3. The circuit must operate in a manner similar to that of a dynamic limiter, and the load resistance R_3 is small to make the diodes damp the secondary circuit L_2C_2 heavily. The time constant R_3C_3 is large and the voltage developed across this network cannot change quickly. The signal voltage across L_2C_2 thus tends to remain constant in spite of any variation in the amplitude of the current supplied to the primary circuit. If this circuit were used purely as an amplitude limiter, the two diodes could, of course, be replaced by a single diode: two are necessary, however, to derive the a.f. output and the mechanism of operation may be explained in the following way.

When the input to the detector is frequency-modulated the voltages applied to the two diodes vary as in a Foster-Seeley circuit: that is to say, if upward frequency swings cause the input to

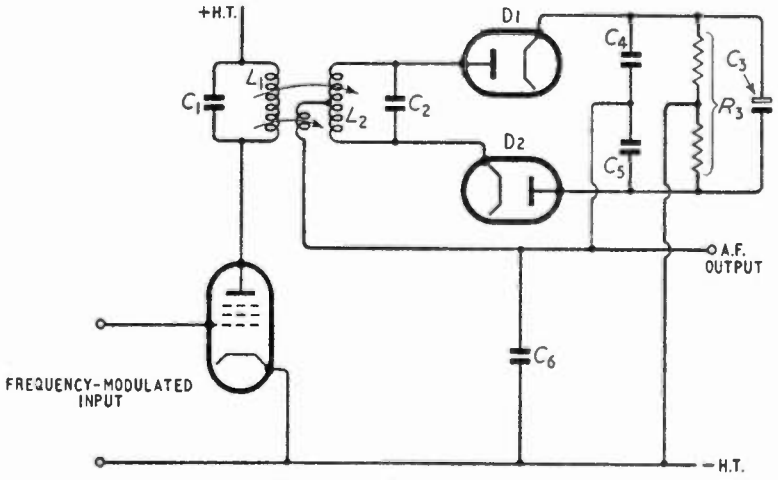


Fig. 5.17. Basic circuit for one type of ratio detector

D1 to increase and that to D2 to decrease, then downward frequency swings will cause the input to D1 to decrease and that to D2 to increase. The diode D1 charges up the capacitor C_4 to a potential approximately equal to the peak input to D1. Similarly, the diode D2 charges up the capacitor C_5 to a potential approximately equal to the peak input to D2.

Thus, the voltages across C_4 and C_5 vary in the same manner as the inputs to D1 and D2. The sum of the voltages applied to D1

and D2 remains substantially constant in spite of frequency modulation and thus the voltage generated across $(C_4 + C_5)$ is also approximately constant. This voltage is also across R_3C_3 and the long time constant of this combination helps to stabilise the voltage at a value proportional to the amplitude of the signal across L_2C_2. The voltage across C_4 (and across C_5) varies in sympathy with the frequency changes, and either capacitor may be taken as a source of a.f. output. If, as shown, the centre point of R_3 is earthed, then for an input current at the centre frequency the voltages across C_4 and C_5 are in fact equal and the potential of the junction point is zero.

For this circuit arrangement therefore the a.f. output swings positively and negatively. If, however, one end of R_3 is earthed, the output from the junction of C_4 and C_5 will be positive or negative at the centre frequency and the a.f. output is at all times positive or negative with respect to earth. C_6 is an r.f. bypass capacitor.

Reduction of Balanced a.m. Component

Although the circuit is superficially similar to the dynamic limiter, its behaviour is not entirely similar, primarily because the ratio detector is a frequency-sensitive circuit. The dynamic limiter operates by varying the damping on a resonant circuit, thus varying the gain of the previous stage so as to maintain constancy in the signal voltage across the tuned circuit. It will follow, therefore, that limiting is achieved by variations in the Q value of the tuned circuit.

In a ratio detector variations in Q of the circuit L_2C_2 can bring about frequency modulation of the signal which can give rise to unwanted output from the detector. If the input current to the primary circuit is at the centre frequency of the detector, variations in the amplitude of the current do not give rise to interference because the detector output is zero for an input at the centre frequency.

If, however, the input current is displaced from the centre frequency, variations in amplitude give rise to an output from the detector, the magnitude of which is proportional to the difference between the input and centre frequencies.

For this reason the output due to this form of interference is frequently referred to as a balanced a.m. component. Fortunately, interference due to this cause can be reduced by modifying the circuit as indicated in Fig. 5.18 so that only a fraction of the diode load is decoupled by the large electrolytic capacitor C_3. The

fraction depends on the design of the detector, but it is commonly of the order of two-thirds so that, if the total diode load is 12,000 ohms, resistor R_3 should be 8,000 ohms, leaving the undecoupled resistor R_4 of 4,000 ohms.

The fraction of the diode load to be stabilised varies with the input-signal amplitude, and the critical value is best determined

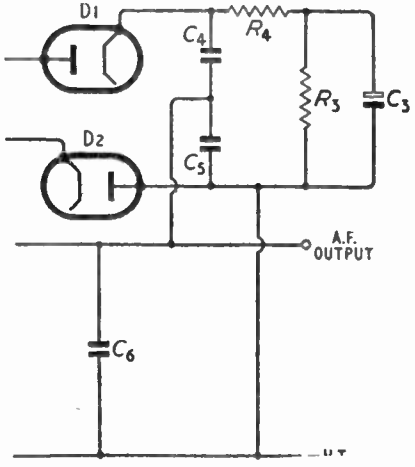


Fig. 5.18. Modification of the circuit of Fig. 5.17 to minimise balanced a.m. component

experimentally when the detector has the amplitude of input signal it is likely to receive under working conditions.

Reduction of Unbalanced a.m. Component

This modification of the basic circuit brings about a substantial improvement in a.m. suppression ratio, but further improvement is possible by minimising a second type of a.m. breakthrough: this gives an a.f. output from the detector which is of constant amplitude and independent of the input frequency, provided this lies within the passband of the detector. This a.f. output is known as an unbalanced component, and it has been attributed to a number of causes, notably variations in diode input capacitance with signal amplitude and inadequate r.f. decoupling of the diode output circuit.

The unbalanced a.m. component can be reduced by the following expedient. The undecoupled resistance R_4 (Fig. 5.18) is made up

of two resistances R_5 and R_6 in series, one being included in diode D1 circuit and the other in diode D2 circuit (as shown in Fig. 5.19), the ratio of R_5 to R_6 being adjusted to minimise the unbalanced a.m. component. During variation of this ratio the total resistance $(R_5 + R_6)$ must be kept constant in order to keep the balanced a.m. component at a minimum.

By adopting these techniques to minimise balanced and unbalanced a.m. components it is possible to obtain a.m. suppression

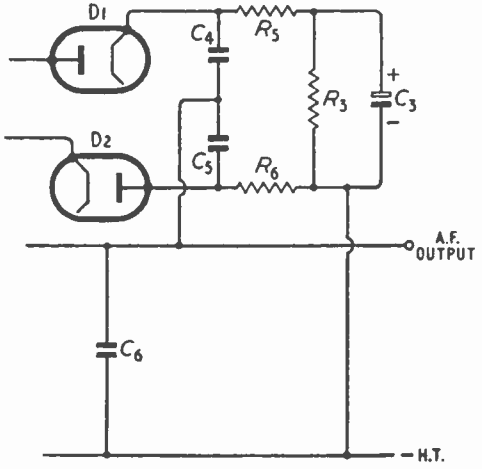


Fig. 5.19. Modification of the circuit of Fig. 5.17 to minimise balanced and unbalanced a.m. components

ratios of 30 dB over a passband of ±60 kc/s. With a performance of this order, most commercial receiver manufacturers dispense with other forms of limiting and operate the i.f. stages at maximum gain to achieve maximum sensitivity.

Provision of Automatic Gain Control

To avoid tuning difficulties the time constant R_3C_3 cannot be increased above approximately 0·1 or 0·2 sec and with such a time constant the ratio detector cannot eliminate slow variations in carrier input (e.g. those having a frequency of 0·5 sec or less, such as those caused by aircraft) although it deals effectively with more rapid fluctuations. To combat slow variations a system of automatic gain control can be used, and a convenient source of a.g.c. voltage is the capacitor C_3 of the ratio detector which can be

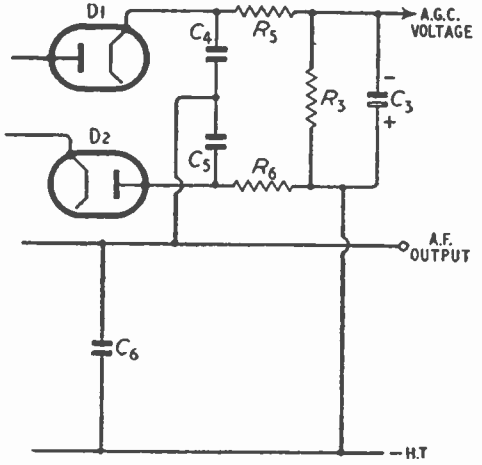


Fig. 5.20. Modification of the circuit of Fig. 5.19 to give an a.g.c. voltage

arranged as shown in Fig. 5.20 to give the required negative voltage for injection into the control grid or the suppressor grid of an i.f. stage.

A.g.c. has another virtue, namely, that it tends to make the a.f. output of the receiver independent of the amplitude of the input signal. Without a.g.c. the ratio detector gives an a.f. output proportional to the amplitude of the input; this follows because the limiting level of the ratio detector is approximately proportional to the amplitude of the input signal.

SELF-LIMITING PHASE-DIFFERENCE DETECTORS

General

There is a further important type of f.m. detector which relies for its action on the phase difference between the signal voltages across two coupled tuned circuits. This detector has a degree of inherent a.m. suppression comparable with that of a ratio detector and gives a much greater a.f. output (30 volts peak is common) but it gives more distortion than the Foster-Seeley or ratio detector circuits.

This detector requires a valve with two input electrodes both of which control the anode current. It is possible to employ the control grid and suppressor grid of a pentode, but this type of valve is not ideally suited for this purpose and a better performance is obtainable from other valves such as the nonode or so-called gated-beam valve which have been specially developed for this purpose.

The anode current must depend on the voltage of each of the control electrodes according to the relationship illustrated in Fig. 5.21. This curve is linear over the greater part of its length but has a nearly-horizontal portion at each extreme. If one of the input electrodes is so biased that the operating point is at the centre of the linear portion, then the anode current is a substantially undistorted copy of small input signals. If the input signal amplitude is increased, a value is reached at which the signal cannot

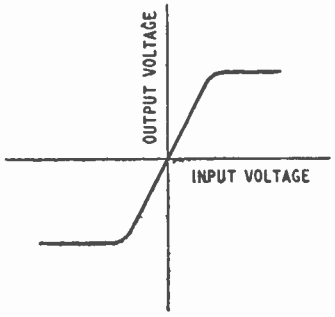


Fig. 5.21. Form of input-output characteristic required in an ideal limiter

be accommodated on the linear portion of the curve. On the crests of the input signal the operating point moves on to the horizontal portions of the curve and further increase in input-signal amplitude fails to produce a corresponding change in anode current. Increase in input signal simply results in the anode current developing into a square waveform with a frequency equal to that of the input signal.

This is the basis of the limiting action of this type of detector and for good limiting the anode current-grid voltage curve must have horizontal sections at the extremes. A conventional control grid does not give a curve of this shape, because every positive increment in the potential applied to the grid always produces some increase in anode current. Even when the control grid-cathode potential is positive and grid current is flowing the anode current still increases if the grid is made more positive.

A better approximation to the ideal limiting characteristic is given by a suppressor grid. This is because the control exercised by a suppressor grid over the cathode current of a pentode is quite different from that exercised by a control grid. In fact the significant point about a suppressor grid is that its voltage does not control the magnitude of the cathode current at all: it controls the ratio in which this current is shared between the two electrodes on

either side of the suppressor grid. In a pentode the suppressor grid is situated between the screen grid and the anode, and the design is usually such that at zero suppressor grid-cathode voltage the anode current is three or four times the screen grid current. In an r.f. pentode, for example, these currents might be 8 mA and 2 mA respectively, giving a total cathode current of 10 mA.

As the suppressor grid is made more negative the anode current decreases and the screen grid current increases, the total remaining constant at 10 mA. At a suppressor grid-cathode potential of, say, −5 volts both currents might be equal at 5 mA, and at a potential of −10 volts the anode current might be zero and the screen grid current 10 mA, as shown in Fig. 5.22. If the suppressor grid is made positive with respect to the cathode the anode current stays at 8 mA and the screen current stays at 2 mA, this ratio of currents measuring the relative intercepting power of the two electrodes.

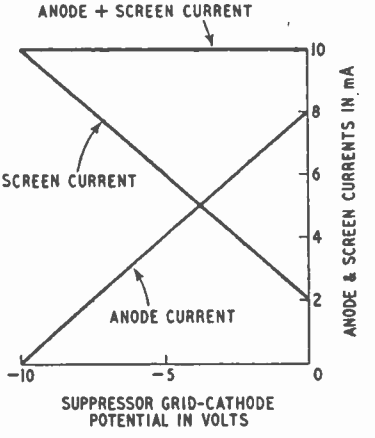


Fig. 5.22. Variation of anode and screen currents with suppressor-grid voltage

Further positive increase in the suppressor grid potential cannot increase the anode current further and limiting is hence good. The only way the cathode current can be increased is to change the control grid or screen grid potentials.

Nonode

Thus, in one form of f.m. detector of this type there is a cathode, control grid and screen grid 1 (all operating at constant potentials to give a constant cathode current), suppressor grid 1 (to which one input is applied), screen grid 2, suppressor grid 2 (to which the

second input is applied), screen grid 3, suppressor grid 3 (held at constant potential to give an overall pentode characteristic) and anode. This makes a total of 9 electrodes, and a valve of this type is thus known as a nonode.

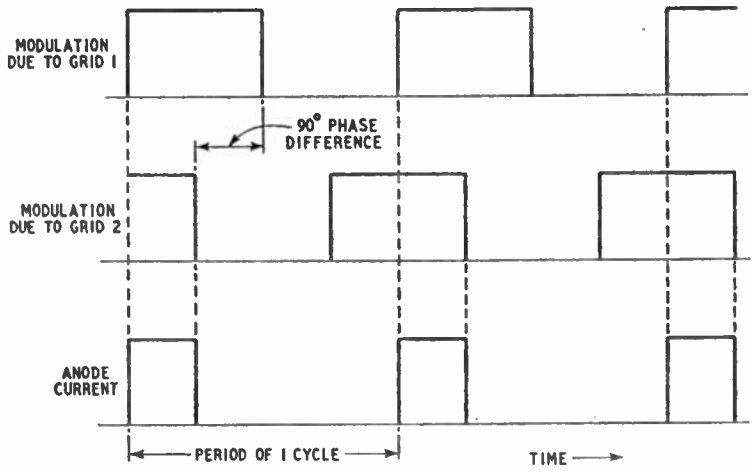


Fig. 5.23. Waveform of anode current of a nonode valve when the two input signals are in quadrature

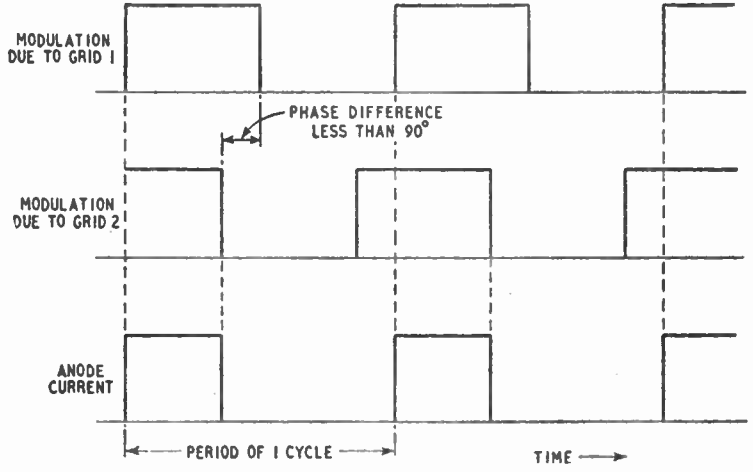


Fig. 5.24. Waveform of anode current when the phase difference between the two input signals is less than 90°

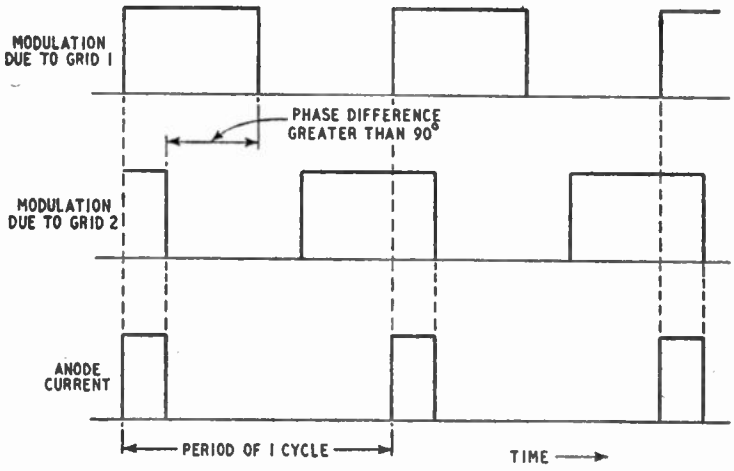


Fig. 5.25. Waveform of anode current when the phase difference between the two input signals is greater than 90°

If the inputs to both grids are well above limiting level, each modulates the anode current by a square waveform, the current alternating between zero and a maximum value. In a detector the signals are obtained from two tuned circuits coupled together, each resonant at the centre frequency of the f.m. transmission. At the centre frequency the signals developed across the two tuned circuits are in quadrature (the reason for this is given in this chapter in the description of the Foster-Seeley circuit) and the anode current is thus modulated by two square waves in quadrature. As shown in Fig. 5.23 this means in effect that the anode current is at its maximum for 90° of each cycle and is at zero for the remainder of each cycle.

When the frequency of the f.m. signal to the detector increases, the phase angle between the two signals applied to the valve changes as indicated in Fig. 5.24, with the result that the anode current pulses increase in width. When the frequency of the f.m. signal decreases, the phase angle between the signals applied to the valve change in the opposite sense, with the result that the anode current pulses decrease in width as shown in Fig. 5.25. Provided the input signals are well above the limiting level, the anode current pulses are of constant amplitude but their width varies with the frequency of the input signal. By including an RC circuit of suitable time constant in the anode circuit of the valve the variations in pulse

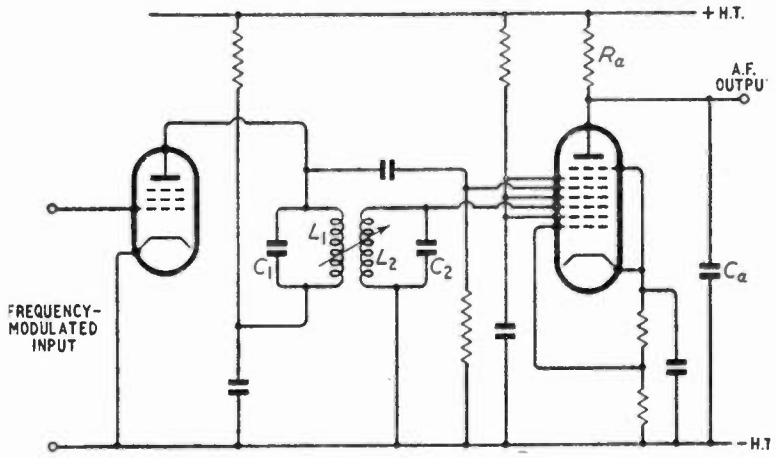


Fig. 5.26. Circuit for a nonode f.m. detector

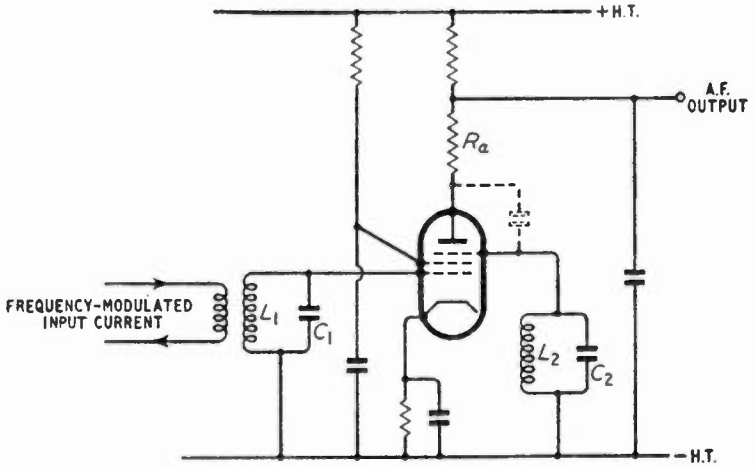


Fig. 5.27. Circuit for a gated-beam f.m. detector

width give rise to a varying voltage which constitutes the a.f. output of the detector. A complete circuit of the detector is given in Fig. 5.26.

Gated-beam Valve

An alternative form of the detector employs the gated-beam valve, and a typical circuit is given in Fig. 5.27. The valve is basically a pentode but, by employing an electron-lens technique in its construction, it is possible to eliminate the disadvantages of the control grid outlined above and obtain good limiting on positive peaks. The two tuned circuits L_1C_1 and L_2C_2 are coupled by means of the electron stream of the valve. The voltage across L_1C_1 modulates the anode current and produces a signal of the same frequency but in antiphase at the anode. This signal is coupled to L_2C_2 by the anode-suppressor grid capacitance, and the magnitude of this signal which controls the degree of coupling between the two circuits can be adjusted by variation of the anode resistor R_a.

6

F.M. RECEIVERS

Introduction

WE shall now discuss the design of domestic f.m. receivers and will then deal in some detail with those sections of such receivers which have not already been described: these include the r.f. stage, oscillator and i.f. stage. A.f. amplifiers and mains-rectifying circuits will not be described because they are not peculiar to f.m. receivers.

GENERAL ASPECTS OF DESIGN

To obtain the requisite gain and passband at the carrier frequencies employed for f.m. broadcasting, it is essential to use a superheterodyne circuit in f.m. receivers.

First consider the passband required. We showed earlier that in a system using a deviation of ± 75 kc/s the bandwidth required to accommodate the significant sidebands is approximately 240 kc/s. A greater passband gives no appreciable improvement in a.f. fidelity but allows some wandering of the oscillator frequency. However, increased bandwidth also causes greater noise voltages to be delivered to the detector and thus impairs the signal-noise ratio. Moreover, increased bandwidth means decreased selectivity and possibly increased interference from signals on neighbouring frequency channels. Thirdly, increased bandwidth inevitably means some reduction in i.f. stage gain and hence in receiver sensitivity.

A value of 240 kc/s is, therefore, to be taken as a maximum rather than a minimum for the i.f. passband and the oscillator must have great frequency stability to keep the i.f. signal within such a passband. By careful design the drift of an oscillator can be kept within 30 kc/s from cold, but for less effective designs automatic frequency correction may be desirable to hold the drift within these limits.

Now consider the gain required. In this country the edge of the service area of an f.m. transmission is taken as the point where the

field strength is 250 μV per metre, measured on a dipole aerial at a height of 30 feet. Signals from indoor aerials may be much smaller—say 10 μV per metre. At the carrier frequencies used for f.m. transmissions a dipole aerial is approximately 2 metres in length and if the signal strength is 10 μV per metre the voltage delivered by a dipole on open circuit is 20 μV. This is halved if, as is usual, the aerial is terminated in a resistance equal to its generator resistance. Thus, the voltage delivered to the receiver input terminals is approximately numerically equal to the field strength, 10 μV in this example.

If the receiver has a Foster-Seeley discriminator preceded by a grid limiter the signal at the limiter grid should, for good limiting, be at least three times the grid base. The grid base can be reduced by using a low screen-grid potential, but if this potential is made too low the pentode becomes incapable of delivering adequate output signal to the discriminator. Thus, there is a lower limit to the screen grid potential and the grid base cannot be decreased much below approximately 1 volt. Thus, for adequate limiting a 3-volt input signal is required and, if the aerial input is taken as 10 μV, the gain required from aerial input to limiter grid is 3×10^5. An i.f. stage is likely to have a gain of approximately 50, and to achieve the required sensitivity two i.f. stages before the limiter are necessary. An r.f. stage is usually considered essential, and thus the receiver may consist of an r.f. stage, a frequency changer, three i.f. stages (of which the third is the limiter), and the discriminator.

Now consider a receiver using a ratio detector. Although limiting does not cease when the input signal to a ratio detector is decreased below a certain threshold value (as in a grid limiter), the performance does, nevertheless, deteriorate with decreasing input and an input signal of preferably not less than 1 volt is required for adequate limiting. If the receiver input is 10 μV, a gain of 10^5 is required: this again needs two i.f. stages and the receiver therefore contains an r.f. stage, frequency changer, two i.f. stages and the ratio detector. Such a receiver contains one valve less than that employing the Foster-Seeley discriminator, and the ratio-detector circuit is generally employed in commercial receivers even though limiting is unlikely to be so effective as that using a grid limiter and Foster-Seeley circuit. By employing a double triode as an r.f. amplifier and a self-oscillating mixer (in a circuit arrangement described later) the number of valves required before the ratio detector can be reduced to three. Nevertheless, this is one more than is customary in medium-wave a.m. receivers

and it implies that f.m. receivers must be more expensive than medium-wave types.

The provision of adequate gain does not in itself guarantee that the full advantages of f.m. will be obtained. We have shown earlier that, in general, a wanted signal can swamp interfering signals provided that the peak value of the carrier exceeds that of the interfering signal. Probably the most troublesome type of interference is that due to the ignition systems of motor cars, and the interfering carrier can give signals of 1 mV peak value on a dipole 30 feet high. To suppress such interference an f.m. receiver must receive a wanted signal exceeding 1 mV and this suggests that the first-class service area of an f.m. transmitter should be regarded as regions where the signal strength exceeds 1 mV per metre.

It does not follow that good results are unobtainable with a lower field strength. It is possible, by using directional aerials, to favour the wanted transmission and/or discriminate against the interfering signal. By this means it may be possible to arrange for the wanted signal to swamp the other. However, to obtain good f.m. reception of a signal of only 10 μV peak value surroundings are required which are very quiet electrically: it may be possible in rural areas remote from a main road.

The ability of an f.m. receiver to suppress impulsive interference often depends critically on the tuning, and to obtain optimum a.m. suppression some form of tuning indicator is virtually essential. This should preferably be a null indicator fed from the output of the detector. The type of indicator often fitted to British f.m. receivers is a magic eye fed from the reservoir capacitor of the ratio detector. The indications given by this type of indicator are not precise. If the i.f. amplifier has a flat-topped response (as it should) the indicator gives a virtually constant indication over a frequency range of 50 to 100 kc/s and over the very region where the detector has zero output and where maximum a.m. suppression occurs. A convenient and inexpensive type of null indicator has yet to be devised.

R.F. AMPLIFIER

An f.m. receiver requires high sensitivity for satisfactory performance, and weak signals may be comparable with valve or circuit noise in early stages. The received signal is weakest in the first stage and to obtain a good signal-noise ratio the first stage should be particularly quiet. It is not good practice, therefore, to have the mixer as the first stage because mixers introduce more noise than

amplifying stages. It is usual therefore to include an r.f. stage before the mixer to raise the level of the input signal above the noise level in the mixer. Even though the gain of an r.f. stage operating at 90 Mc/s is quite low, it improves the signal-noise ratio provided that it does not itself introduce considerable noise. The inclusion of an r.f. stage also necessitates a further signal-frequency tuned circuit; this is useful because it improves signal-frequency selectivity and reduces second-channel interference.

Thirdly, an r.f. stage behaves as a buffer between the oscillator and the aerial and reduces radiation of oscillator output from the aerial. Radiation from the oscillator of superheterodyne v.h.f. receivers can be appreciable and can cause serious interference on

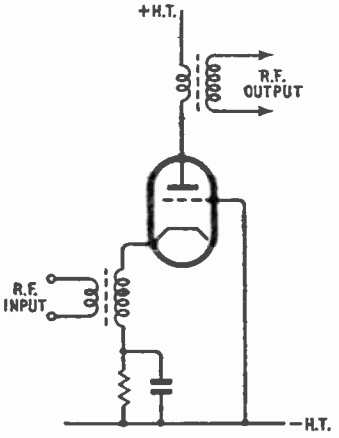


Fig. 6.1. Essential features of earthed-grid r.f. amplifier

other v.h.f. receivers, and receivers must be specifically designed to minimise such radiation.

The r.f. amplifier can be a pentode, but is more usually a triode. There is little to choose between the two in terms of gain but the triode gives less noise because of the absence of partition noise. If the triode is connected as a conventional earthed-cathode amplifier, it requires neutralising to avoid instability when the grid and anode circuits are tuned to the same frequency.

This can be avoided if the triode is used as an earthed-grid amplifier as in the circuit of Fig. 6.1. Here the control grid behaves as a screen between output and input circuits and helps materially in reducing radiation of oscillator output from the aerial.

The input resistance of an earthed-grid triode is $1/g_m$, only

200 ohms if g_m is 5 mA/V. The dynamic resistance of the tuned circuit at 90 Mc/s is unlikely to be higher than a few thousand ohms but, nevertheless, the damping imposed by the earthed-grid triode is so heavy that there is little point in attempting to tune the circuit by varying the inductance or capacitance. Usually the aerial circuit is fixed-tuned to the centre of the band (say to 94 Mc/s); the loss in signal transfer from aerial to cathode at the extremes of the band (87·5 and 100 Mc/s) is negligible.

It might be thought that the input resistance of an earthed-cathode triode would be much higher than that of the earthed-grid circuit but the difference is not as marked as might be supposed. There are a number of sources of damping, such as transit-time effects, cathode-lead inductance and feedback from the output circuit to the input circuit via the anode-grid capacitance. In a practical triode these sources might give an input resistance of the order of 2,000 ohms at 90 Mc/s.

Because of the low input resistance of r.f. stages, the aerial coupling circuit must be designed to match the characteristic impedance of the aerial feeder to the input resistance of the valve. For example, if the cable has 80 ohms resistance and the valve has an input resistance of 2,000 ohms the turns ratio of the aerial transformer is given by

$$\sqrt{(2{,}000/80)} : 1 = \sqrt{25} : 1$$
$$= 5 : 1$$

If the tuning inductor has 5 turns, the aerial winding requires only a single turn.

MIXER

The purpose of a mixer stage is to obtain from the received signal and from the local-oscillator output a resultant with a frequency equal to the difference between the frequencies of the two input signals. In general there are two basic principles which can be used to achieve this.

In one method the two input signals are simply connected in series or in parallel and applied to the input of a valve. If the anode current-grid voltage relationship for the valve is linear, the two signals are amplified independently and there is no output at any frequency other than those of the two input signals. In order to produce inter-action between the two input signals which is essential to give an output at the difference frequency the valve must have a non-linear characteristic and should thus behave as a

detector. For this reason the mixers of early superheterodyne receivers were known as first detectors. Stages of this type are known as *additive mixers* and they are usually biased near the point of anode-current cut-off to achieve the non-linearity essential for their action.

In other more modern types of mixer the received signal and the local oscillator output are in effect multiplied together. This is achieved by applying the two inputs to separate electrodes of a valve which is so designed that a signal applied to one electrode controls the gain of a signal applied to the other electrode. Two such electrodes are the control grid and the suppressor grid of a pentode. In a mixer of this type there is no need for any non-linearity, the output at the difference frequency being produced directly from the effective multiplication of the two input signals. Such mixers are termed *multiplicative* and most modern mixers such as hexodes or heptodes are of this type. Clearly it is wrong to refer to such mixers as first detectors because they can be linear devices.

The additive type of frequency changer is preferred in f.m. receivers because it offers a higher conversion conductance (e.g. 2 mA/V) than multiplicative types (0·7 mA/V) whilst its noise is less. The mixer may be a triode or a pentode and a typical circuit is given in Fig. 6.2. The signal-frequency input is applied by choke-capacitance coupling from the r.f. stage, and the oscillator

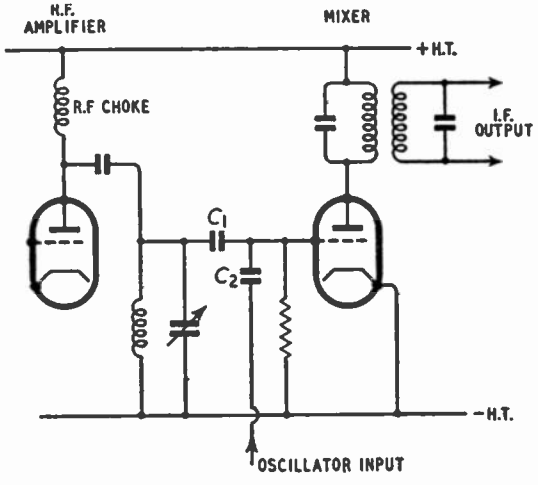


Fig. 6.2. Essential features of an additive mixer stage

input is applied to the grid in parallel with the signal-frequency input via a small series capacitor.

In the circuit illustrated the mixer has no fixed bias and obtains its negative potential by rectification of the oscillator input in the diode formed by the grid and cathode. Grid current flows in positive half-cycles of the input signal from the oscillator and charges up the capacitor C_2. The conversion conductance of the mixer, i.e. the ratio of difference-frequency current output to signal-frequency voltage input, increases with increase in oscillator drive and reaches maximum at a particular oscillator input, after which the conductance slowly falls as indicated in Fig. 6.3. In

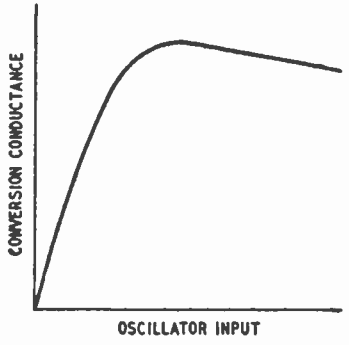


Fig. 6.3. Variation of conversion conductance with oscillator input for a mixer stage

addition the damping imposed by the mixer grid circuit on the signal-frequency input decreases as the oscillator drive is increased and maximum gain from the r.f. stage is achieved with an oscillator input slightly greater than that giving maximum conversion conductance. This adjustment also ensures that slight variations in oscillator drive (which might occur when the oscillator tuning is altered) do not cause significant changes in mixer gain.

In another version of this circuit the mixer valve is biased by a resistor in the cathode circuit and a smaller oscillator drive is employed. To obtain maximum gain from the mixer under such conditions, the cathode resistor must be chosen to bias the valve near a point where the $I_a - V_g$ characteristic has marked curvature. This circuit has the advantage that only a small oscillator input is required, and this in turn permits greater frequency stability from the oscillator: this point is discussed more fully later.

In a mixer of the type illustrated in Fig. 6.2 the mixer signal-frequency circuit must have variable tuning and this is normally ganged with the oscillator tuning, the resonance frequencies of the two circuits differing by the intermediate frequency at all settings of the tuning control. If the oscillator frequency is above the signal frequency, the mixer circuit behaves as a capacitance at the oscillator frequency, and this capacitance forms with C_2 a capacitive potential divider which attenuates the oscillator signal applied to the mixer.

Provided the intermediate frequency is small compared with the signal and the oscillator frequencies, the attenuation introduced by this divider does not vary greatly with tuning and the oscillator can be designed to inject optimum voltage into the mixer grid in spite of the loss in the divider.

If the oscillator frequency is below the signal frequency, the signal-frequency circuit is effectively inductive at the oscillator frequency. The inductance does not greatly vary with tuning, provided the intermediate frequency is small compared with the signal and oscillator frequencies. Thus C_2 forms with the signal-frequency circuit a CL potential divider and the attenuation introduced by this decreases with increase in frequency. Thus, the variations of oscillator drive with this circuit are inevitably greater than those obtained in a circuit in which the oscillator frequency is above the signal frequency. The variations will not cause significant changes in receiver sensitivity if the oscillator drive is adjusted to be greater than that which is required for optimum conversion conductance.

If the signal-frequency circuit at the mixer input is fixed-tuned, the resonance frequency is most likely to be placed near the centre of the band to be received. Irrespective of whether the oscillator frequency is above or below the signal frequency, the difference between the oscillator frequency and that to which the mixer circuit is tuned must be small when the receiver tuning approaches one end of the band. At such a setting the mixer circuit ceases to be predominantly reactive (as when the mixer circuit has variable tuning) and acquires a resistive component which can impose serious damping on the oscillator. This is sometimes heavy enough to prevent oscillation altogether at one end of the waveband and causes a serious loss in sensitivity as the tuning is advanced towards that end.

The decision whether the oscillator frequency should be above or below the signal frequency is usually settled by considerations

of the interference likely to be caused to other receivers by oscillator re-radiation. It seems that the interference is likely to be least when the oscillator frequency is above the signal frequency, and this is the mode of operation which is favoured in most commercial receivers.

OSCILLATOR

The oscillator in an f.m. receiver may be a Colpitts, a Hartley or a Reinartz type, but the preference is usually for the Colpitts type and a typical circuit is given in Fig. 6.4. Probably the most serious design problem is that of securing adequate frequency stability.

The stability is considered good if the frequency does not vary by more than 30 kc/s. At 100 Mc/s this represents a frequency change of only 0·03 per cent and is not easy to achieve.

The chief cause of frequency drift is the change in valve input capacitance as the valve warms up, although drift is also caused by variations in the inductance and the capacitance in the oscillator circuit, as these become warmer due to increase in the temperature

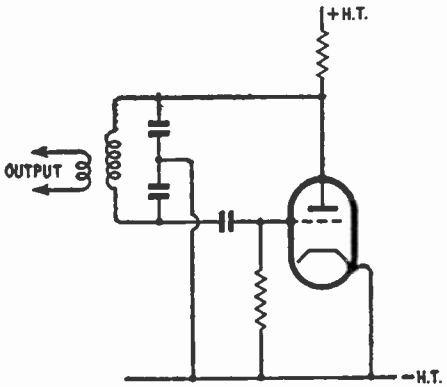


Fig. 6.4. Colpitts oscillator

of the receiver as a whole. The effect of changes in valve input capacitance can be reduced by using a circuit in which the input capacitance is " tapped down " the frequency-determining LC circuit and so has reduced effect on oscillator frequency: such a circuit is illustrated in Fig. 6.5. If, however, the valve is tapped too far down the tuned circuit, the degree of regeneration becomes

insufficient to maintain oscillation and the oscillator ceases to operate.

This sets a lower limit to the extent to which the valve can be tapped down.

In general, increase in temperature causes the capacitance of capacitors (including that of a valve input circuit) and the inductance of inductors to increase, thus lowering the frequency of oscillation. Capacitors can, however, be so constructed that the

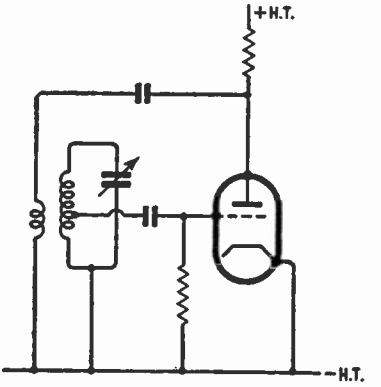


Fig. 6.5. Illustrating one method of "tapping down": the oscillator is a Reinartz type

capacitance decreases with increase in temperature, and negative coefficients which are as large as 2,200 parts in 10^6 per °C can be achieved.

By including such capacitors in the tuned circuits of v.h.f. oscillators it is possible to offset the increase in inductance and of valve input capacitance by the decrease of capacitance of the negative temperature coefficient capacitors, and by this means a highly stable oscillator can be produced.

So successful can be this process of compensation that some of the commercial f.m. receivers now available have no observable tuning drift at all.

SELF-OSCILLATING MIXERS

In the design of commercial receivers one of the aims is always to reduce the number of valves to a minimum. One way in which economy can be effected is to use a single triode as a self-oscillating

mixer, and a typical circuit is given in Fig. 6.6. The circuit is simply that of a Colpitts oscillator, with provision for injecting a signal-frequency input and for abstracting an intermediate-frequency output. Care must be taken in the choice of the input injection point to minimise feedback of oscillator output to the r.f. stage. One of the techniques commonly employed for this purpose can be explained in the following way.

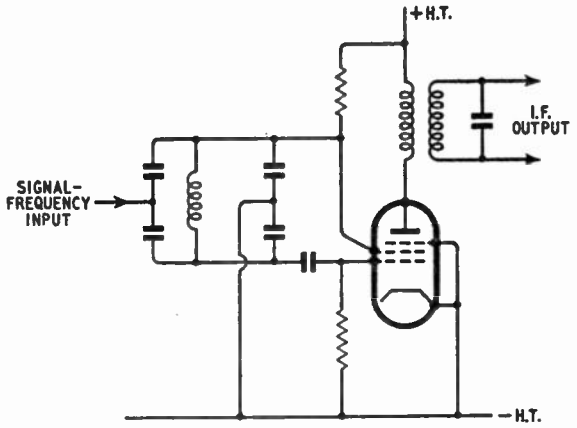


Fig. 6.6. One type of self-oscillating mixer using a pentode

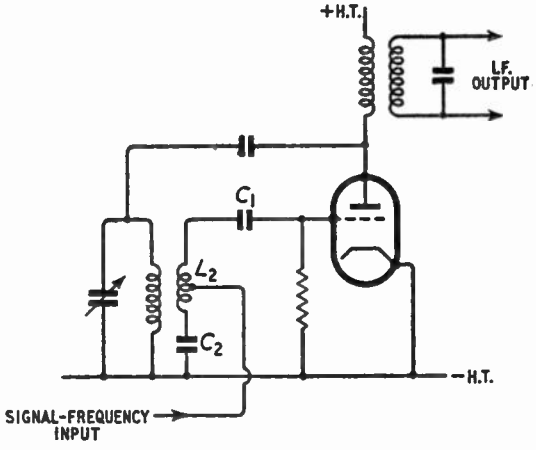


Fig. 6.7. One form of self-oscillating mixer using a triode

Consider a Colpitts oscillator in which the LC frequency-determining circuit is connected between the anode and the grid of the maintaining valve, with one capacitor connected from anode to cathode and another from grid to cathode, the cathode being at earth potential. In such an oscillator both ends of the inductor are at r.f. potential, but the potential at the anode is in antiphase to that at the grid and there is a point on the inductor at which the r.f. potential is zero. This point is effectively at earth potential and is known as a null point. If the two capacitors are equal, the null point is at the centre of the inductor. If two further capacitors are connected in series across the inductor so as to form a capacitance divider, the junction point can be made a null point by choosing the values of the capacitors appropriately. This is the basic principle of the circuit of Fig. 6.6.

A null point can be found in those types of Hartley oscillator and in other types of oscillator for which the r.f. potential at one point of the inductor is in antiphase with respect to that at another point. The null point is a convenient point at which to inject a signal-frequency input because the oscillator output at that point is a minimum and a number of self-oscillating mixers make use of such circuits. One example is given in Fig. 6.7: here the lower half of L_2 and capacitor C_2 are chosen to have equal reactances at the signal-frequency. The reactances are opposite in sign, but a common current flows in both and it follows that the signal-frequency voltage generated across the two components are equal in amplitude but opposite in phase. Since one plate of C_2 is earthed, the tapping point of L_2 must also be.

At the signal frequency the null-point input impedance is capacitive and can be tuned to resonance at the signal frequency by suitable choice of inductor in the generator feeding the null point. It is desirable that the oscillator output at the null point should remain at a minimum whilst the oscillator frequency is varied during tuning: the method of tuning must be chosen with care to achieve this. For example, in a Colpitts oscillator one of the two fundamental capacitors must not be varied to effect tuning because these capacitors determine the position of the null point on the inductor or on any potential divider connected across the inductor. Tuning must therefore be effected by variation of both capacitors simultaneously or by variation of the inductance.

Triodes are useful as mixers but have the disadvantage that their anode a.c. resistance is often low enough to act as a serious shunt on the primary winding of the i.f. transformer connected in

the anode circuit. To maintain high gain this damping can be reduced by application of positive feedback to the mixer, and one circuit used for this purpose is illustrated in Fig. 6.8. Feedback is applied via the components R_1 and C_2 and the feedback circuit may be regarded as approximating to that of a Colpitts oscillator in which the degree of regeneration is deliberately made too small for oscillation. The degree of regeneration is controlled by the ratio of C_1 to C_2, these acting as the two fundamental capacitors of the Colpitts oscillator.

In commercial receivers it is common practice to use a double-triode in the early stages, one acting as an r.f. amplifier and the

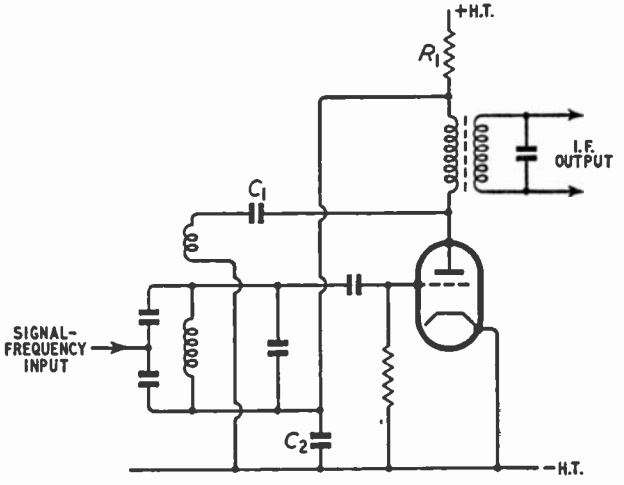


Fig. 6.8. One method of eliminating damping of the primary winding of the i.f. transformer by positive feedback

other as a self-oscillating mixer. Thus, all the v.h.f. circuits are associated with a single valve and often all the v.h.f. components are accommodated within a metal screening box on which the double-triode is mounted.

VALVE ARRANGEMENT

A few commercial receivers are designed for f.m. reception only, but most can receive a.m. transmissions on medium waves and long waves in addition. As pointed out earlier, the circuit commonly used for f.m. reception includes one r.f. stage, a frequency

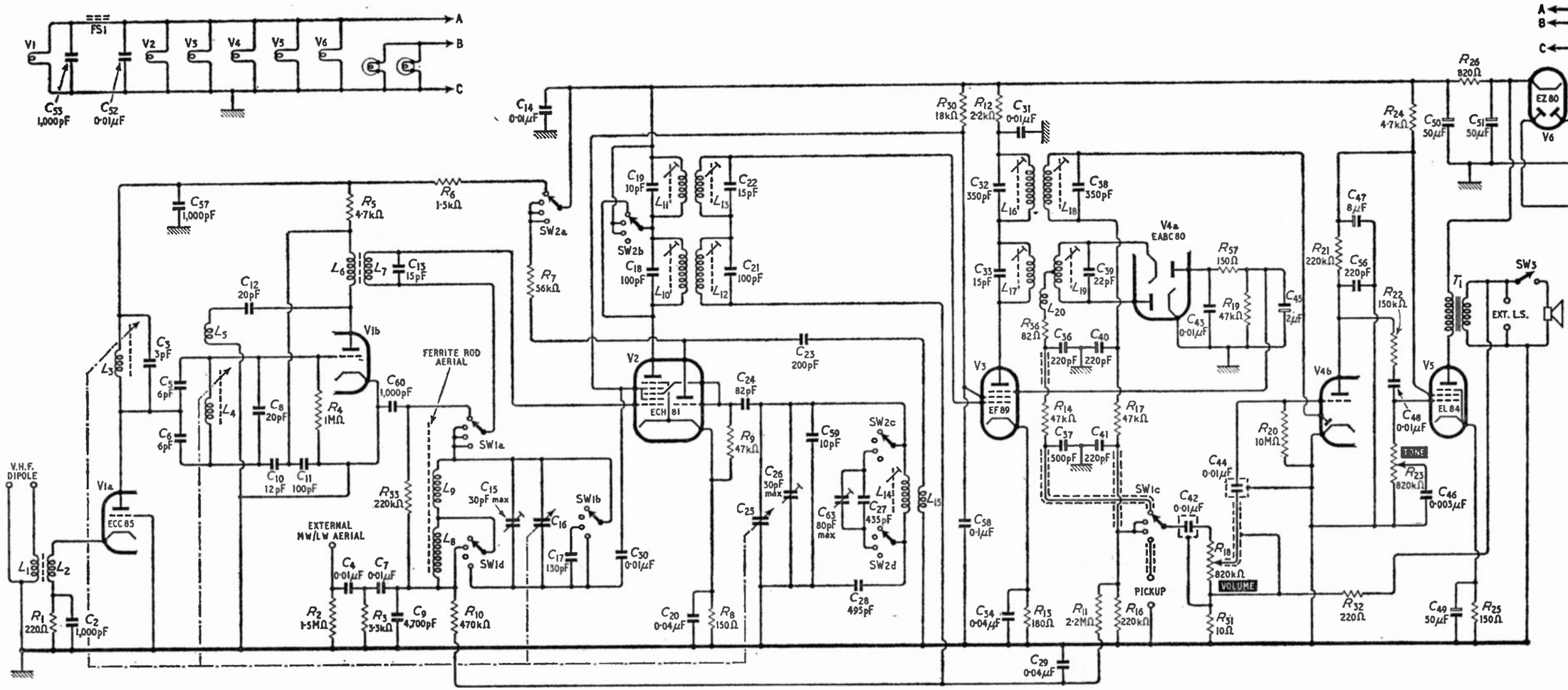


Fig. 6.10. Circuit diagram of a commercial f.m.–a.m. receiver

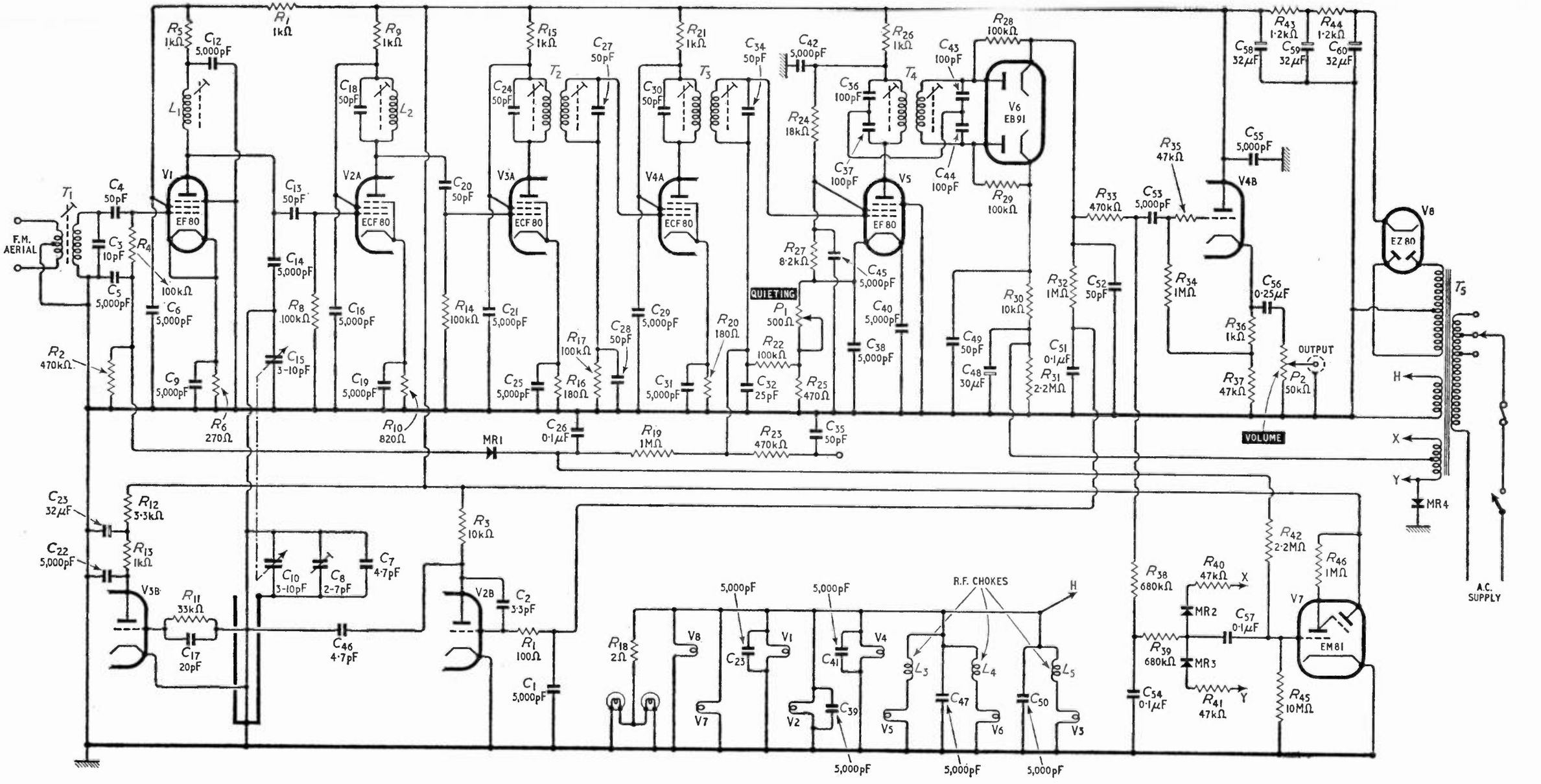


Fig. 6.11. Circuit diagram of a commercial f.m. tuner

changer, and two i.f. stages preceding a ratio detector. For a.m. reception a frequency changer and a single i.f. stage before the detector are adequate. The problem is how to obtain both circuits, using reasonably simple switching and without a multiplicity of valves. One solution is indicated in Fig. 6.9. All the pre-detector stages are achieved with only three valves, namely a double-triode,

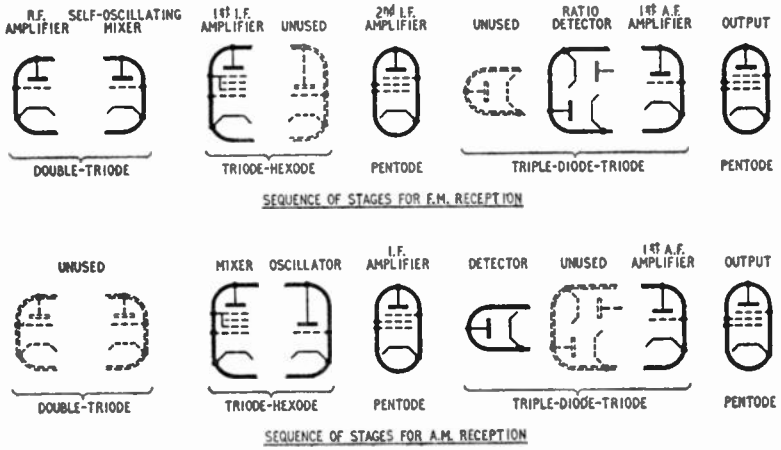


Fig. 6.9. One method of using five multi-purpose valves in an f.m.-a.m. receiver

a triode-hexode and a pentode. For f.m. reception the double-triode behaves as r.f. amplifier and self-oscillating mixer, the hexode acts as first i.f. amplifier (the triode being unused) and the pentode as second i.f. amplifier.

For a.m. reception the double-triode is unused; the triode-hexode is used as mixer and oscillator with the pentode as the only i.f. amplifier. A triple-diode-triode has been developed for use in such receivers, two diodes being used in the ratio detector, the third diode as a.m. detector and the triode as first a.f. amplifier. A pentode output stage completes the receiver which employs only five valves (exclusive of rectifier) in spite of its complexity.

F.M.-A.M. RECEIVER

The complete circuit diagram for one commercial f.m.-a.m. receiver of this type is given in Fig. 6.10 (opposite page 112). The waveband switch is shown in the position for f.m. reception and

the following is a description of the circuit arrangement on this waveband. The feeder from the internal or external v.h.f. aerial leads to an r.f. transformer L_1L_2 which feeds into the cathode circuit of triode V1A, which is one half of a double valve V1. The damping at the cathode circuit is approximately $1/g_m$, 200 ohms for a g_m of 5 mA/V. This constitutes heavy damping and there is no advantage in having variable tuning in this transformer. It is therefore designed to be resonant at the centre of Band II. The anode circuit of V1A is tuned by adjustment of the inductor L_3, the fixed capacitance comprising the output capacitance of V1A, C_3 and the input capacitance of the self-oscillating mixer V1B.

V1B is designed as a Reinartz oscillator, regeneration occurring via L_4 and L_5. The frequency is varied for tuning by adjustment of the inductance L_4, the core of which is ganged with that of L_3. Oscillator re-radiation is minimised by injecting the r.f. input into the null point of the oscillator circuit, this point being obtained by the capacitors C_5 and C_6 across the circuit. L_6 and L_7 constitute the first i.f. transformer and damping of the primary circuit is avoided as explained earlier, by positive feedback due to R_5 and the two capacitors C_{11} and C_{12}.

The hexode section of V2 is an i.f. amplifier operating at a centre frequency of 10·7 Mc/s and feeding the second i.f. transformer $L_{11}L_{13}$. The pentode V3 is the second i.f. amplifier and feeds the third i.f. transformer $L_{17}L_{19}L_{20}$. This is the ratio-detector transformer and feeds the double diode V4A. One plate of the reservoir capacitor C_{45} is earthed and the other gives a negative output which is proportional to the input-signal amplitude and is returned to the suppressor grid of V3 to give a.g.c. The a.f. output from the ratio detector is taken from the winding L_{20}, is decoupled by $R_{36}C_{36}$ and then passes via the de-emphasis network $R_{14}C_{37}$ to the volume control R_{18}.

V4B is a triode first a.f. amplifier which is biased by the 10-MΩ grid resistor R_{20} and is RC-coupled to the output pentode V5. A negative feedback voltage is taken from the secondary winding of the output transformer and is injected into the bottom end of the volume control. In this way feedback is made to vary with the volume-control setting, being a maximum at low settings (i.e. on strong signals) and a minimum at high settings (i.e. for weak signals). The mains rectifying and smoothing circuit is conventional.

When the receiver is switched for medium- or long-wave reception or for gramophone reproduction, wafer SW2A breaks the h.t.

supply to V1 and the windings L_8 or L_9 on the Ferrite rod are used as aerial. Provision is also made, however, for using an external aerial. This is connected via the network $R_2C_4R_3C_7$ across C_9 which is in series with the signal-frequency tuned circuit. A two-gang variable capacitor is used for medium- and long-wave tuning, section C_{16} being included in the signal-frequency circuit and C_{25} in the oscillator circuit. An unusual feature of the circuit is the method of obtaining long-wave oscillator tuning. Instead of using a separate long-wave oscillator inductor, additional capacitors $C_{27}C_{63}$ are connected in parallel with the medium-wave inductor L_{14}. A.g.c. is obtained from the load resistor R_{16} of the diode in V4B which is also used as a signal detector.

F.M. TUNER

To illustrate some of the points made earlier we shall now examine the circuit diagram of a high-fidelity f.m. tuner: this is given in Fig. 6.11, opposite page 113. The unit is completely self-contained, incorporating its own mains rectifying equipment, and employs a total of eight valves, three of which are triode-pentodes.

The first stage is an r.f. pentode V1 with tuned grid and tuned anode circuits. A.g.c. bias is supplied via the rectifier MR1 as explained later. V2A is a pentode mixer stage with a single tuned circuit L_2C_{18} resonant at the intermediate (centre) frequency in the anode circuit. The oscillator input is injected via the 1-pF capacitor C_{11} to V2A grid. The oscillator circuit employs a trough line instead of the more usual inductor. This is a length of transmission line, the outer conductor of which is of square cross-section but has one side missing. The centre conductor is a rod to which connections are made through the gap which is left by the missing side.

The line is short-circuited at one end and its length is such that it presents an inductive reactance at the tapping points on the centre conductor, to which the grid and cathode of the oscillator valve V3B are connected. The line is tuned by the variable capacitor C_{10} at its open-circuited end and thus the oscillator valve is effectively tapped down the oscillating circuit. V2B is a reactance valve connected across the line to reduce any drift that may occur in oscillator frequency.

The fundamental components of the reactance valve are the capacitor C_2 between anode and grid (augmented, of course, by the inter-electrode capacitance) and the resistor R_1 between grid

and cathode (C_1 is a decoupling capacitor of low reactance). R_1 is returned to the discriminator output.

V3A and V4B are i.f. amplifiers and the components $R_{17}C_{28}$ are included to enable V4B to operate as an additional limiter for strong signal inputs. V5 is, however, the principal limiter, which operates with a low (50 V) screen potential. A d.c. bias is developed across R_{22} by grid-current flow and this bias is used for a.g.c. purposes, being returned to V1 grid via R_{19} and MR1. R_{22} is returned to a tapping point on V5 cathode and thus the a.g.c. voltage is positive in the absence of an input signal. The rectifier MR1 is included to prevent this bias being applied to V1. MR1 does not conduct until the negative voltage which is developed across R_{22} by the i.f. signal at V5 grid exceeds the positive bias across R_{25}.

Thus there is no a.g.c. action until the input to the tuner exceeds a certain threshold value. The variable resistor P_1 is included in V5 cathode circuit to act as a quieting control. By advancing this, V5 can be completely cut off and the receiver then gives no noise output in the absence of an input signal. Strong signals then stand out against a completely quiet background.

The discriminator transformer T_4 has a capacitive centre tap on primary and secondary windings and feeds the double diode V6.

The a.f. output is developed across C_{52}. A.f. signals are attenuated by R_{33} and C_{51} to give a d.c. output for the reactance valve. This voltage swings positively and negatively with respect to the cathode potential of the lower diode of V6 (depending on the sign and the magnitude of any oscillator mistuning). V2B, however, has an earthed cathode (to avoid negative feedback and consequent loss of reactance-valve sensitivity) and thus the a.f.c. voltage must at all times be negative to prevent V2B being driven into grid current. To achieve this V6 cathode is returned to the negative voltage across C_{48}, this voltage being obtained from the heater supply by the rectifier MR4 which is effectively connected across half the heater winding of the mains transformer.

The discriminator output is also fed to the cathode follower V4B, the cathode circuit of which includes the volume control P_2. The output is also fed to the magic eye tuning indicator V7 which operates in the following manner. A.f. signals are eliminated by $R_{38}C_{54}$ to give a d.c. output which is dependent on oscillator mistuning.

This output is zero (with respect to the potential at V6 lower

cathode) for exact tuning. A voltage of this value can be short-circuited without effect on its value. Thus at the correct tuning point the input to V7 can be short-circuited without effect on the light pattern of the magic eye. Short circuits of the input occur 50 times per second and are achieved by rectifiers MR2 and MR3 connected across the heater winding. If the oscillator is mistuned there is a d.c. input to V7 and short-circuiting of this input gives an alteration of the light pattern. Thus the eye gives a blurred indication if there is mistuning and a crisp indication for accurate tuning. The indication is also crisp, however, even in the absence of an input signal (for which the discriminator output is also zero).

This ambiguity is eliminated by feeding the a.g.c. voltage also into V7 grid via the resistor R_{47}. The biasing of V7 is such that the sectors of luminescence are scarcely visible when there is no signal input to the receiver, but become visible as signals are tuned in, the display being blurred except at the correct tuning point.

AERIAL

Commercial receivers usually incorporate internal a.m.-f.m. aerials and these are quite effective on strong signals. A ferrite rod aerial is generally employed for a.m. signals and a compressed dipole

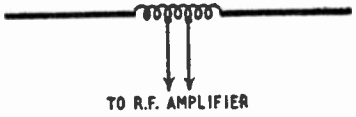


Fig. 6.12. One method of loading a simple dipole to secure resonance at a frequency less than that of natural resonance

for f.m. reception. The f.m. aerial commonly consists of two short lengths of wire or two strips of foil arranged horizontally in the cabinet. In a small receiver these conductors cannot, of course, be the full length (approximately 5 feet) required for resonance at 90 Mc/s but resonance can be achieved by loading shorter lengths of conductor with inductors as suggested in Fig. 6.12. Such aerials give good f.m. reception in situations where the wanted signal is greater in peak value than interfering signals.

If the receiver is used at ground level near a main road, reception free of ignition interference may be possible only within a few miles

of the f.m. transmitter where the wanted signal is strong enough to over-ride the interference. At higher situations, e.g. at first or second floor level where the distance from main roads is greater, interference-free reception is possible at greater distances from the transmitter. If the receiver must be used close to a road and at some distance from the transmitter, it may be essential to use a full-sized dipole in the loft or on the roof and connected to the receiver by a length of feeder. Receivers are usually provided with sockets for the connection of such an aerial.

7

NON-BROADCASTING APPLICATIONS OF FREQUENCY MODULATION

APPLICATION OF FREQUENCY MODULATION IN MICROWAVE COMMUNICATION SYSTEMS

Introduction

IN previous chapters we have discussed methods of frequency modulating carrier waves up to 100 Mc/s and have described the techniques used in transmitters and receivers for such frequencies. 100 Mc/s is not, of course, the highest radio frequency used for communications purposes. Frequencies higher than 300 Mc/s and up to, say, 7,000 Mc/s are extensively used for communicating over short, quasi-optical distances. These frequencies correspond to wavelengths of less than 1 metre and are known as microwave frequencies. Such waves cannot be generated or amplified by conventional valves and specialised types such as klystrons have been developed for these purposes. These valves lend themselves well to frequency modulation and in this section we shall describe the principles of such valves and of the circuits associated with them.

Limitations of Conventional Valves as Frequency Increases

The performance of conventional valves such as r.f. pentodes deteriorates as frequency is raised. The deterioration is due to a number of losses which become more important as frequency is increased. The principal sources of loss are the following.

Transit-time effects

When the time taken for an electron to travel from the cathode of a valve to the control grid becomes comparable with the periodic time ($1/f$) of the signal applied to the grid, the valve draws appreciable power from the signal source. This power can be represented as a resistance connected between control grid and cathode, and the

value of this resistance is approximately inversely proportional to the square of the frequency. This can be shown in the following way.

Consider an alternating voltage given by

$$v = v_0 \sin \omega t$$

applied between the grid and the cathode of a valve. If the frequency $\omega/2\pi$ of this voltage is low, the time taken for electrons to cross the gap between the cathode and the grid is a very small fraction of the periodic time of the signal and may be neglected in comparison with it. We can then say that the density of the electron stream arriving at the grid is in phase with the applied signal. The electron stream impresses upon the grid an electric charge which is also in phase with the applied signal and can thus be represented by

$$q = q_0 \sin \omega t$$

Now the current flowing between the grid and the cathode is given by the rate of change of the charge with respect to time and is thus given by

$$i = \frac{dq}{dt}$$

$$= \frac{d}{dt}(q_0 \sin \omega t)$$

$$= \omega q_0 \cos \omega t$$

This current thus leads the applied voltage by 90 degrees, a property of a capacitance. Thus the grid-cathode of a valve behaves as a pure capacitance at low frequencies.

Now let the frequency of the applied signal be raised to a value for which the transit time is an appreciable fraction of the periodic time. The alternation of electric charge now lags that of the applied voltage by an angle ϕ where

$$\phi = \omega T$$

and T is the transit time. For an applied voltage given by $v = v_0 \sin \omega t$, the induced charge is now given by

$$q = q_0 \sin (\omega t - \phi)$$

and the current between the grid and the cathode is given by

$$i = \frac{dq}{dt}$$

$$= \frac{d}{dt}\{q_0 \sin(\omega t - \phi)\}$$

$$= \omega q_0 \cos(\omega t - \phi)$$

$$= \omega q_0 \cos \omega t \cos \phi + \omega q_0 \sin \omega t \sin \phi$$

For a given valve and applied signal ω, q_0 and ϕ are all constant and this expression shows that the grid-cathode current now has two components. The first leads the applied voltage by 90 degrees as at low frequencies, indicating that the grid-cathode impedance has a capacitive component. The second component of the current is, however, in phase with the applied signal and therefore represents a resistive component of input impedance in parallel with the capacitive component.

The value of the resistance is given by the voltage divided by the current, i.e.

$$R = \frac{v_0 \sin \omega t}{\omega q_0 \sin \omega t \sin \phi}$$

$$= \frac{v_0}{\omega q_0 \sin \phi}$$

Now q_0 is proportional to v_0 and thus the resistive component is proportional to $1/\omega \sin \phi$. For small values of ϕ, $\sin \phi$ is approximately equal to ϕ which, in turn, is equal to ωT. Thus the resistive component of impedance is proportional to $1/\omega^2 T$ and, for a given transit time, is inversely proportional to the square of the frequency.

For an r.f. pentode the resistance may be as low as 5,000 ohms at 50 Mc/s and hence 1,250 ohms at 100 Mc/s. Such low values of input resistance seriously limit the stage gain obtainable from the valve.

Effects of Stray Inductance and Capacitance

The values of inductance and capacitance used in tuned amplifiers must be reduced to give resonance at higher frequencies.

For example, a medium-wave tuned circuit may have an inductance of 180 μH and a capacitance of 200 pF. A circuit to resonate at 45 Mc/s may have an inductance of 1 μH and a capacitance of 15 pF and in circuits resonating at this and at higher frequencies it is quite usual to employ stray capacitance or the input capacitance of a valve as the tuning element. To obtain resonance at much higher frequencies requires so small a capacitance and so small an inductance that it is quite likely that the stray capacitance may be too large and the inductance of even short connecting leads may be too large.

It is clear, therefore, that there is an upper limit to the resonance frequency which is obtainable with conventional valves and components.

Effects of Power Losses

Losses in dielectrics, due to skin effect and to radiation from conductors all tend to increase as frequency is raised.

Up to approximately 200 Mc/s the performance of conventional valves is generally adequate but between this frequency and 1,000 Mc/s the gain falls to unity and it becomes necessary to use specialised valves to obtain a useful performance.

Disk-seal Valves

These valves do not lend themselves as well to frequency modulation as do the reflex klystrons described later but are, nevertheless, briefly described here to give a reasonably comprehensive survey of the types of valves employed up to 7,000 Mc/s.

Disk-seal valves may be used up to approximately 2,500 Mc/s. These are valves basically of conventional form but in which the construction has been specifically designed to minimise the effects tending to degrade high-frequency performance. For example, transit-time effects are minimised by reducing the control grid-cathode spacing to a very low value. Stray capacitance and inductance is reduced by shaping the electrodes in the form of disks which project through the glass envelope and are flanged at the outer edge to locate with the co-axial lines which are used as frequency-determining elements.

Fig. 7.1 illustrates an r.f. amplifier which employs a disk-seal triode.

The output circuit consists of a length of concentric line, the conductors of which locate with the anode and grid disks of the triode.

The input circuit similarly employs a length of concentric line the cylinders of which will locate with the control grid and cathode disks.

For convenience a common cylinder is employed as the outer conductor of the input circuit and also as the inner conductor of

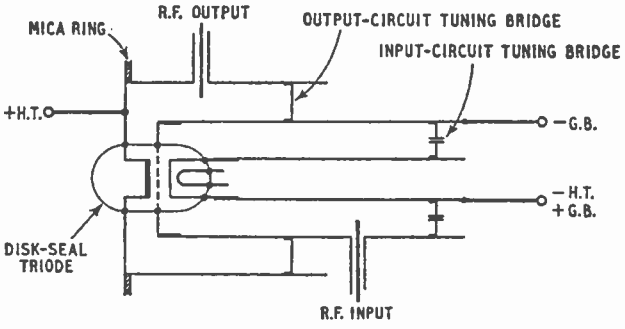


Fig. 7.1. A disk-seal triode used as an r.f. amplifier

the output circuit: thus this amplifier is an example of a common-grid type.

The lengths of the concentric lines are adjusted to give an inductive input reactance which resonates with the capacitive reactance of the valve input or output circuit at the frequency of operation. The input and output connections to the amplifier can be made by co-axial feeders which terminate in probes which enter the concentric lines as illustrated.

By reducing the grid-cathode spacing of a conventional type of valve to the absolute minimum as in disk-seal valves, worthwhile gain at, say, 3,000 Mc/s can be obtained but valves operating on entirely different principles are necessary to achieve reasonable gains at higher frequencies.

Typical of such valves are klystrons which can be used as r.f. amplifiers and generators up to 7,000 Mc/s. They employ a principle known as velocity modulation which makes use of transit time in order to achieve amplification.

Velocity Modulation

Consider Fig. 7.2 which represents a valve with a heater, a cathode, four grids and an anode. The grids are arranged as two

pairs with a large space between them. When the l.t. and h.t. supplies are connected electrons are released from the cathode and travel to the anode, passing through the grids on their way. They reach the anode, under the attraction of its positive charge, as a stream of uniform density. The electrons take an appreciable time to travel between the cathode and the anode and thus cross the valve with a particular value of velocity. Now suppose an r.f. voltage is applied between grids 1 and 2, the pair nearer the cathode. The voltage varies sinusoidally with time and produces an electric field which is parallel to the electron path and also varies sinusoidally with time. This field accelerates or retards the electrons

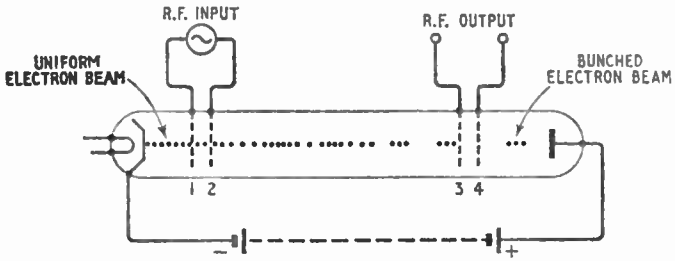


Fig. 7.2. Illustrating velocity modulation

depending on its direction and amplitude, and the velocities of the electrons leaving grid 2 is no longer constant but has a component varying sinusoidally with time. The velocity of the electrons has been modulated and the process is termed velocity modulation.

Two-cavity Klystron

In the space between the pairs of grids the varying velocities of the electrons affect their spatial distribution because those with velocities greater than average tend to overtake those which were released a little earlier from the cathode but move more slowly. Similarly, those with velocities less than average tend to lag behind and get closer to those which were released from the cathode a little earlier but move more quickly. Thus, the electrons tend to gather into bunches but this takes a little time and the bunches are more clearly defined at some distance from grid 2 than close to it.

There is, in fact, a particular distance from grid 2 at which the bunches are most clearly defined and this distance depends on the anode-cathode voltage and on the amplitude of the r.f. voltage

applied to grids 1 and 2. Grids 3 and 4 are arranged to be at this critical distance and the bunches of electrons, in passing between these grids, induce voltages between them which have the same frequency as the input signal. These voltages constitute the output of the valve which is thus taken from a pair of grids and not from

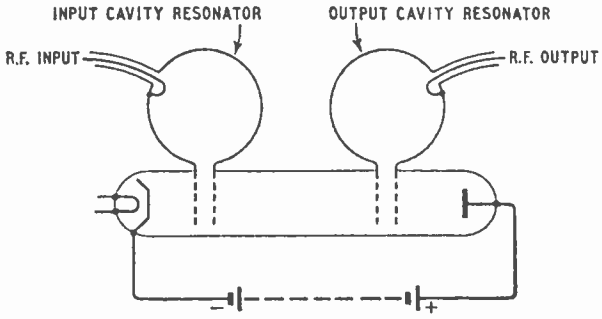


Fig. 7.3. Basic form of two-cavity klystron

the anode as in a conventional valve. This is illustrated in Fig. 7.2. The anode in a velocity-modulated valve functions purely as a collector of electrons and plays no part in signal amplification.

If grids 3 and 4 are connected to a circuit which resonates at the frequency of the input signal, an amplified output signal can be obtained. This is the principle of many microwave valves. The resonant circuit cannot be a conventional LC type or even a length of concentric transmission line at very high frequencies and it is commonly a closed conducting container known as a cavity resonator, the dimensions of which are carefully chosen to suit the frequency of operation. The resonator may be regarded as an LC circuit because it has capacitance (between opposite walls, for example) and inductance (the walls form continuous loops of conductor). Usually the input grids are also connected to such a cavity resonator and thus the basic form of one type of velocity-modulated valve is as shown at Fig. 7.3. The input signal is introduced by means of a co-axial feeder, the inner conductor of which forms a small loop in the input resonator and the output signal is taken away also by means of co-axial feeder, the inner of which also forms a loop in the output resonator.

This diagram represents the essential features of the two-cavity klystron. It may be compared with a conventional amplifier with a tuned input and a tuned output circuit. If the two cavities

have the same resonance frequency the valve may be used as a narrow-band amplifier. If the two cavities are coupled together (e.g. by means of a length of co-axial feeder) the klystron may be used as an oscillator.

Reflex Klystron

To generate oscillations in a klystron it is not necessary to employ two resonators. A single resonator may be used, its two grids producing velocity modulation as described above. The electron beam can then traverse a path long enough to give good bunching and can then be returned to the same pair of grids. Provided the signal induced in the resonator by the return beam is in the right phase and of sufficient amplitude, regeneration and consequent oscillation can occur. A klystron employing this mode of operation is known as a reflex klystron and its essential features are illustrated in Fig. 7.4.

The cavity resonator is biased positively with respect to the cathode to attract electrons from the cathode. These pass through the

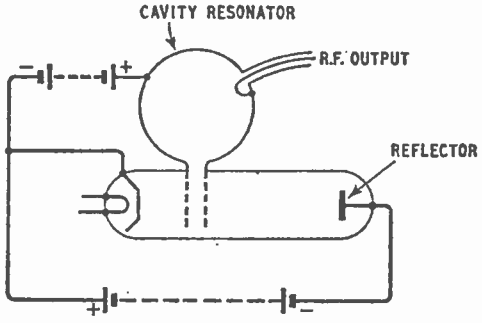


Fig. 7.4. Basic form of reflex klystron

two grids, being velocity modulated as they do so. The electrons then approach a reflector which is negatively biased with respect to the cathode. The decelerating field between resonator and reflector halts the electrons before they reach the reflector and accelerates them back to the resonator, where they again pass through the two grids and are finally collected by the resonator. The reversal in the direction of the electrons does not affect bunching and the operating conditions can be adjusted to give well-defined bunches in the return beam when it reaches the resonator.

These are the conditions necessary to give oscillation and an output can be taken from the resonator by a co-axial cable and loop as illustrated.

Oscillation occurs at a frequency not far removed from the resonance value for the cavity resonator.

Now suppose, in a reflex klystron which is oscillating as just described, the reflector potential is made more negative. This

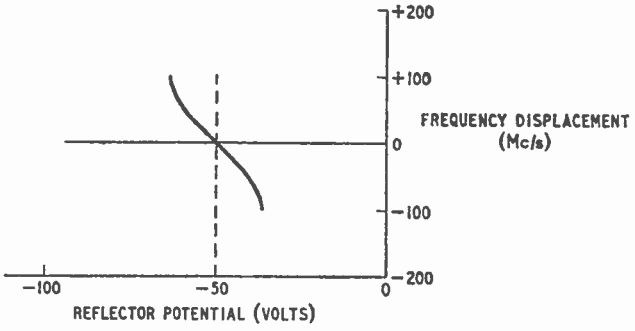


Fig. 7.5. Variation of oscillation frequency with reflector potential in a reflex klystron

makes the retarding field between resonator and reflector stronger and the velocity-modulated electrons are not able to penetrate it as deeply as before. They are, therefore, returned to the resonator more quickly than before and the early arrival of the bunches at the resonator causes an increase in the frequency of the field in the resonator and hence in the oscillator output. Similarly, if the reflector is made less negative, electrons penetrate the retarding field to a greater extent and the bunched beam returns late to the resonator, causing the oscillation frequency to fall. Thus a change in the reflector potential causes a change in the oscillation frequency and the relationship between these two quantities is of the form illustrated in Fig. 7.5: over a small range of reflector potential the curve is substantially linear, making the reflex klystron suitable for frequency modulation. The modulating signal is connected in series with the battery applying the negative bias to the reflector, the battery voltage being adjusted to secure operation on the linear part of the curve.

The sensitivity of the modulator, i.e. the ratio of the frequency change produced, to the modulating voltage producing it, is

determined by the extent to which the resonator is damped by the external load applied to it. Heavy loading gives high sensitivity but it is not desirable to operate the modulator with such loading because the output power varies considerably with the modulating signal as shown in Fig. 7.6, giving unwanted amplitude modulation.

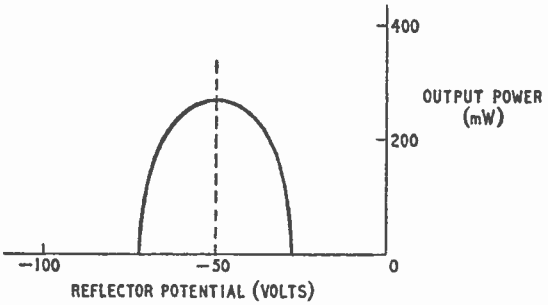


Fig. 7.6. Variation of output power of reflex klystron with reflector potential

For lighter loading the variation of output power is much less and this condition is normally used even though the sensitivity of the modulator is smaller.

The variation of output frequency and output power illustrated in Figs. 7.5 and 7.6 might occur between reflector potentials of, say, −40 and −60 volts, the steady bias being adjusted to −50 volts for linear modulation. If the steady bias is increased beyond −60 volts or reduced below −40 volts, oscillation ceases because bunches now arrive back at the resonator in the wrong phase or with insufficient amplitude to maintain oscillation. However, further increase or decrease of bias will produce other ranges in which oscillation occurs again.

These ranges might be from −75 to −105 volts and from −30 to −15 volts and these ranges of reflector potential can also be employed for frequency modulation if desired although the mean frequency of oscillation is approximately the same for all ranges of bias potential.

To alter the mean frequency of oscillation for a reflex klystron it is usual to alter the volume of the cavity resonator by physical deformation. In one type of klystron one face of the resonator is corrugated and the centre can be displaced against the stiffness of a

return spring by a screw mechanism which is used as a coarse mean-frequency control. Fine adjustment of mean frequency is usually achieved by variation of the reflector bias. The range of the coarse control is commonly about 10 per cent of the mid-range frequency.

For example, the control might cover the range 4,750 Mc/s to 5,250 Mc/s.

Microwave Transmitter

Fig. 7.7 gives the circuit of a reflex klystron frequency modulator. The output of the klystron is fed via a length of co-axial feeder to a waveguide and from thence to the aerial which is commonly a paraboloid.

The depth of penetration of the probe into the waveguide and the position of the waveguide piston relative to the probe are both

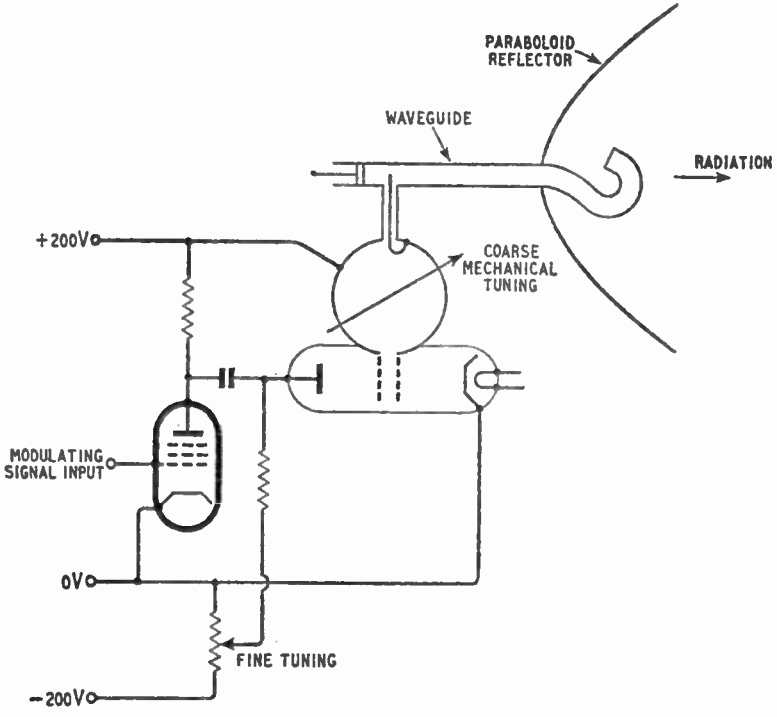


Fig. 7.7. Simplified circuit for reflex klystron frequency-modulated transmitter

adjusted in order to secure maximum transmission of power to the aerial.

The power output of a small reflex klystron may be less than 1 W but the gain of the paraboloid aerial is so large that the effective radiated power may be as much as 10 kW. Such a simple transmitter can give satisfactory communication over distances up to 30 miles or more.

Microwave Receiver

Fig. 7.8 illustrates the essential features of a microwave superheterodyne receiver of the type which can be used in conjunction with the transmitter just described. It employs a paraboloid receiving aerial, a dish-shaped structure 3 or 4 feet in diameter mounted vertically and which must be positioned accurately to obtain the maximum signal from the very narrow beam radiated from the transmitting aerial. (Similarly, of course, the transmitting aerial must be aimed with great accuracy at the receiving site: this is essentially a point-to-point communications system.) The received signal is conveyed along a waveguide and is mixed with the output of the receiver oscillator which is similarly conveyed by a waveguide. For a short distance the two waveguides share a common wall and mixing then occurs by way of an aperture in this wall.

The oscillator is a reflex klystron which may be of the same type employed in the transmitter. It is coupled to its waveguide by a length of co-axial cable terminating in a probe as described earlier. The mixer is a crystal diode (point-contact type) mounted in the aerial waveguide and its output is fed to the i.f. amplifier which commonly operates at a midband frequency of around 50 Mc/s. The discriminator at the end of the i.f. amplifier gives the required modulation-frequency output and may also be used for automatic frequency control of the oscillator.

The oscillation frequency of the klystron can be adjusted roughly to the correct value by mechanical distortion of the cavity and can then be adjusted accurately by the desired value (indicated by zero d.c. output from the discriminator) by adjustment of the reflector bias, a.f.c. voltage being switched off during this adjustment. The a.f.c. switch is then operated to the "a.f.c. on" position.

Some wandering of the oscillator frequency of the transmitter klystron and the receiver klystron is inevitable, even when both valves are enclosed within thermostatically-controlled containers,

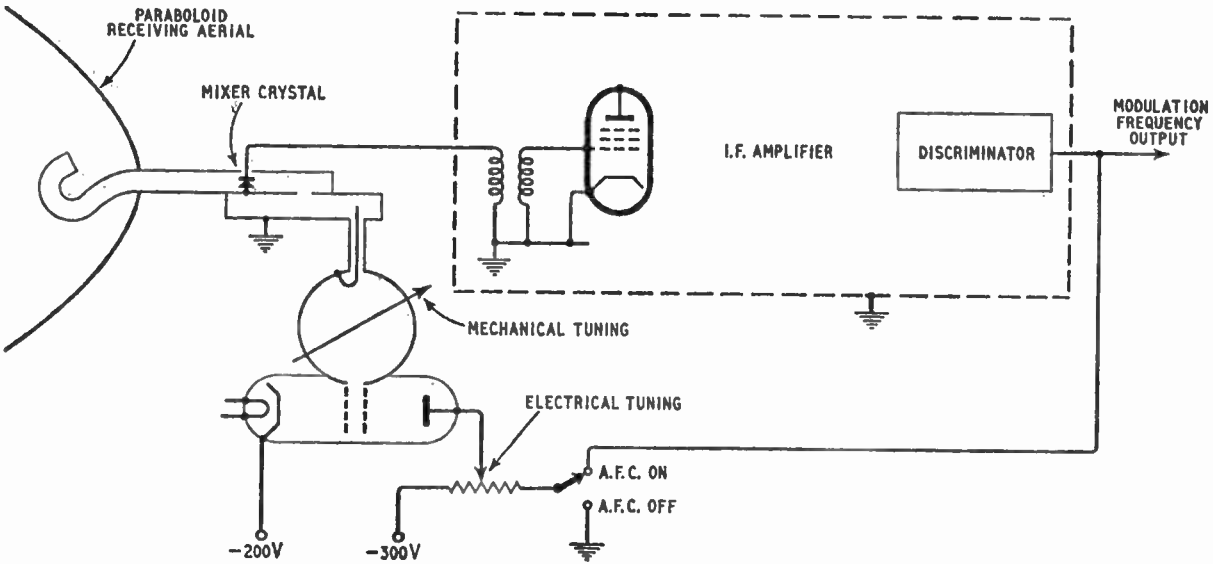


Fig. 7.8. Essential features of a microwave superheterodyne receiver with reflex klystron as oscillator and a crystal mixer

and it is essential to control the frequency of the receiver klystron by an a.f.c. circuit in order to ensure that the mixer output remains within the passband of the i.f. amplifier.

APPLICATION OF FREQUENCY MODULATION IN RADAR

Introduction

To conclude this book we shall deal with some applications of frequency-modulated waves which have not hitherto been mentioned. In previous chapters we have described the nature of frequency modulation and have discussed its use in communications systems, notably in high-quality sound broadcasting. Today frequency modulation plays an important part in other branches of radio, in particular in radar: in this section we shall describe the most important of these uses of frequency modulation.

Radar is the name given to electronic equipment used to detect and locate objects at such great distances or under such conditions that the human eye cannot be used. The equipment can measure the range of the object, its altitude or its speed with great accuracy. It can also provide certain information about the object; for example, it can give a rough outline of the object and even show large details on it. The term "radar" is derived from the initial letters of radio detection and ranging.

Originally radars were designed for military purposes, e.g. for the detection of hostile aircraft and to give information of their distance, height, speed, etc. which is needed for anti-aircraft gun-laying. Radars were also developed during the second world war to give a map of the area beneath a plane: this enabled bombers to locate their target with accuracy. Since the war radars have been employed in a large number of different peace-time activities. They are used as navigational aids by ships and aircraft enabling neighbouring objects to be seen when visibility is bad due to darkness or fog. They are also used in harbours and airports to help in guiding ships and aircraft. In meteorology they are used to trace the movements of balloons and clouds and as storm detectors. They help scientific work, notably astronomy, by tracking meteors and man-made satellites. Possibly it has been most recently applied by the police, who now use radar to measure the speed of cars.

Pulse Radar

Nearly all radars employ the following principle: a radio wave is radiated from a transmitter. Some of the energy is reflected

from the object under investigation and it is returned to the equipment to be picked up by a receiver. The returned energy can be used to give information about the object or can be used to give a picture of it. A number of different types of transmission are used to find out different items of information. In the simplest radar system information about the distance (i.e. range) of the object is obtained by timing radio waves in their journey to the object and back to the equipment. This cannot be achieved with a continuous-wave output from the transmitter. The wave must be modulated in some way, and one of the principal methods of modulation is by interrupting the transmission so as to send out narrow pulses with long intervals between them: this is an example of amplitude modulation.

The time taken for any given pulse to make the double journey can be measured with accuracy and from this the distance of the object can be determined.

Radio waves travel in free space at 186,000 miles per second. This is equal to 0·186 miles per microsecond. Thus, a delay of 10 microseconds between transmission and reception of a pulse means that the wave has travelled 1·86 miles and the object is therefore 0·93 miles away.

This numerical example shows that for short ranges the time of travel of the radio wave is only a few microseconds. The reflected signal must be received as a pulse which does not overlap the transmitted pulse which gave rise to it: thus the pulse duration must be small and is commonly between 0·1 and 10 μsec.

Each transmitted pulse must be followed by a long period of no transmission during which reflected pulses can be received. The duration of the interval between successive pulses depends on the range of the most distant object it is intended to detect. For a range of 20 miles the interval must be at least 200 μsec (corresponding to a pulse repetition frequency of 5 kc/s) and for 100 miles range the interval must be 1 msec (corresponding to 1 kc/s) but intervals of 5 msec (corresponding to 200 c/s) have been used. Thus typical values of the pulse duration and pulse interval are 1 μsec and 1 msec, giving a mark-to-space ratio of 1 : 1,000.

The timing of the pulses permits the range to be determined. To determine the direction of the object it is usual to radiate a very narrow beam from the transmitting aerial. A search of a particular area can then be made by systematic movements of the aerial which cause the beam to scan that area. To give a

narrow beam the dimensions of the aerial must be large compared with the wavelength yet the aerial must be reasonably small because it must be moved during scanning. Thus the radiated wavelength must be small and in practice is usually between 1 cm and 50 cm (600 Mc/s to 30,000 Mc/s). It is usual in radars to transmit and receive with the same highly-directional aerial

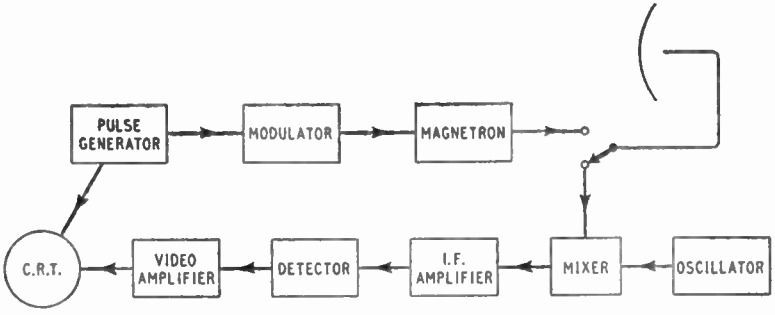


Fig. 7.9. Block diagram of a pulse radar

because of the difficulty of keeping a bulky receiving aerial accurately in step with the transmitting aerial during the scanning process.

A block diagram of a radar is given in Fig. 7.9. A generator producing pulses of the required duration and frequency feeds a modulator which controls an oscillator operating at the carrier frequency. For high carrier frequencies the oscillator is often a magnetron. The oscillator feeds the directional aerial via a transmission line or waveguide. The reflected signals are collected by the same aerial and are applied by a line or waveguide to a mixer, commonly a silicon crystal diode. The local oscillator may be a klystron, the frequency of which is adjusted to give an i.f. output from the mixer of between 30 Mc/s and 100 Mc/s. This is amplified in an i.f. amplifier and then applied to a detector. The pulses received from the detector are further amplified in a video amplifier and are then displayed on the screen of a cathode ray tube. The display is such that it enables the range of the object to be accurately determined. A number of different forms of display are used: two of those most commonly used are described later.

A cell is included in the waveguide, coupling the aerial to the magnetron and a second cell is included between the aerial and the mixer. In its unenergised state, the cell between aerial and

magnetron is closed so as to block this path, whereas the cell between aerial and mixer (in its unenergised state) is open. Thus in this state any received signals are directed into the receiver and cannot enter the magnetron. When the magnetron oscillates, however, the power released into the waveguide energises both cells, thus opening the path between magnetron and aerial but closing the path between magnetron and mixer. In this way radiation from the aerial can take place but the crystal is not exposed to the full output of the magnetron which would normally be sufficient to damage it. Nevertheless a little of the magnetron output does reach the mixer and this is useful because it is normal to display the pulses at the moment of transmission as well as at the moment of reception. The action of the two cells is represented in the block diagram Fig. 7.9 as a simple changeover switch.

In one form of cathode ray tube display the transmitted and received pulses are used for Y (vertical) deflection of the electron beam, a time base providing X (horizontal) deflection. The horizontal distance between the two pulses is proportional to the time taken for the radio wave to travel to the reflecting object and back again to the radar. This, in turn, is proportional to the range of the object and the cathode ray tube screen can be calibrated so that the range can be read off directly.

In another type of display the received pulse is used to control the intensity of the electron beam and each reflected signal is thus represented as a bright patch on the screen. For a single reflection there are two such patches, one produced by the pulse at the moment of transmission and the other at the moment of reception. An advantage of this type of display (known as the PPI, the Plan Position Indicator) is that it can be used to give direction as well as the range of reflecting objects. The line joining the two bright patches is caused to rotate about the patch representing the transmitted pulse in phase with the rotation of the aerial about its vertical axis. In this way the radar screen gives a scale map of the area surrounding the radar in which large objects such as hills and buildings show up as stationary bright patches. Moving objects show up as moving spots of light and their distance from the central bright patch measures the range of the object. The angle of the line joining the spot of light to the screen centre gives the direction of the object. Thus the outer perimeter of the screen can be calibrated to give the bearing of the object relative to the radar.

The radio waves reflected from such objects as hills and large

buildings are used in certain types of radar to give information about these objects. In radars intended for tracking moving objects, for example aircraft, permanent reflections from nearby stationary objects are unwanted because they tend to obscure the reflections from moving objects. This difficulty can be eliminated by use of Doppler radar: this can be regarded as a form of f.m. radar because the reflected signal has a frequency different from that of the transmitted signal, the frequency change occurring as a result of the velocity of the moving object.

Doppler Radar

If a carrier wave of constant frequency f_0 is reflected from an aircraft or other moving object, some of the reflected energy may be picked up by the receiver at the radar. If the object is moving away from or towards the radar the reflected wave is not of the same frequency as the transmitted wave: this is because of an effect known as the Doppler effect.

The reflected wave is, in fact, being radiated from a moving object and this affects the frequency as measured at a stationary point in the following manner.

First suppose the reflecting object is stationary and reflecting waves of frequency f_0. The space between the transmitter and the object and back again to the receiver (this is a distance equal to twice the distance between the object and the radar) is occupied by oscillations having a wavelength λ which is related to the velocity of propagation c by the expression

$$f_0 = \frac{c}{\lambda}$$

These oscillations move through space at the velocity c and for every oscillation leaving the transmitter one is received back at the radar. In one second f_0 cycles are sent out from the transmitter and f_0 cycles are picked up at the receiver.

Now suppose the object is moving directly towards the radar at a velocity v, still reflecting waves of frequency f_0. The total distance from transmitter to object and back again from object to the receiver is now shortening by a distance equal to $2v$ every second and, as this distance is filled with oscillations of wavelength λ, more waves are picked up by the receiver in one second than are sent out by the transmitter. Thus, the received frequency is higher than the transmitted frequency and the difference in frequency is equal to the

number of wavelengths in a distance $2v$. Thus, the received frequency f_r is given by

$$f_r = f_0 + \frac{2v}{\lambda}$$

which may be written

$$f_r = f_0 + \frac{2v}{c} \cdot \frac{c}{\lambda}$$

$$= f_0 + \frac{2v}{c} \cdot f_0$$

$$= f_0\left(1 + \frac{2v}{c}\right)$$

The increase in frequency is given by

$$\text{frequency change} = \frac{2v}{\lambda}$$

Similarly, if the object is moving directly away from the radar at a constant velocity v, the frequency of the reflected signal picked up by the receiver is less than the transmitted frequency, the difference being given by the same expression $2v/\lambda$.

To obtain an estimate of the frequency changes likely to be experienced in practice, we can modify this expression to permit values of v in miles per hour and values of λ in centimetres to be substituted directly. We know that 60 miles per hour is equal to 88 feet per second and that 1 foot equals 12 inches and therefore $12 \times 2{\cdot}54$ centimetres. Thus, v (in miles per hour) must be multiplied by $(88 \times 12 \times 2{\cdot}54)/60$ to give the velocity in centimetres per second. This is equal to 45 approximately and thus we have

$$\text{frequency change} = 90\frac{v}{\lambda} \text{ c/s}$$

As a numerical example suppose v is 300 miles per hour and λ is 10 centimetres. The frequency change is given by

$$\text{frequency change} = 2 \times 45 \times \frac{300}{10} \text{ c/s}$$

$$= 2{\cdot}7 \text{ kc/s.}$$

We have so far been considering objects moving directly towards or directly away from the radar. If the object is moving in a

direction making an angle θ with the straight line joining the object to the radar as shown in Fig. 7.10, then the Doppler radar measures the component of velocity $v \cos \theta$ along the straight line: this is usually known as the radial component.

It should be noted that a Doppler radar measures *radial velocity* not *range*. It can, however, be used to determine the range provided that the radiated wave is modulated in some way. The most convenient method is by amplitude-modulating the transmitted waves by pulses as in normal pulse radar. The range can then be determined by cathode ray tube display as just described and the velocity of the moving object by measuring the difference between the received and transmitted frequencies. This difference

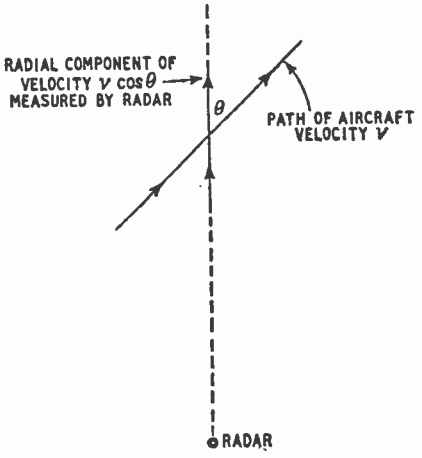


Fig. 7.10. Illustrating radial component of velocity measured by radar

cannot be determined directly because, as shown in the numerical example just given, the changes in frequency are very small. They are therefore determined indirectly by a beating process illustrated in the typical schematic circuit shown in Fig. 7.11.

A local oscillator is employed as in superheterodyne receivers to change the very low difference frequency to a much higher value which can be effectively amplified in an i.f. amplifier. The output of the detector following the i.f. amplifier is an amplified replica of the difference frequency.

In a pulse Doppler radar, signals reflected from stationary

objects have the same frequency as the transmitted signals, whereas signals reflected from moving objects have frequencies slightly higher or slightly lower than that of the transmitted signal. In a display in which Y deflection is proportional to pulse amplitude and X deflection is proportional to range it may not be confusing to have permanent and moving echoes displayed together. In a

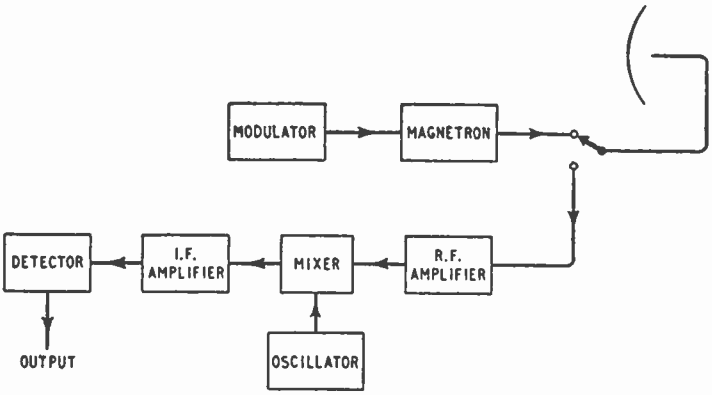


Fig. 7.11. Block diagram of a Doppler radar

PPI display, however, permanent echoes would certainly obscure moving ones and some means of suppressing permanent echoes is required.

One method of achieving this is to delay all received signals by an interval corresponding to the pulse recurrence frequency and to subtract the delayed from the undelayed signals. Permanent echoes are of constant amplitude and occur at the transmitted pulse recurrence frequency. Thus, by suitable adjustment of the amplitude, the delayed and the undelayed signals can be made to cancel. Signals reflected from moving objects have an amplitude and a pulse recurrence frequency which vary with time and they do not necessarily therefore cancel in the subtraction circuit.

The problem then is to delay pulses of say 1 μsec duration by an interval of say 1 msec without significant change of waveform. An electrical network would be inconveniently complex for this purpose and an ultrasonic delay cell containing water or other liquid is commonly used instead. The velocity of sound in such cells is round 4,500 feet per second and a cell 4½ feet long is thus

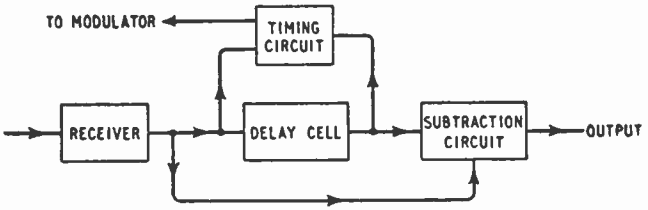


Fig. 7.12. Block diagram illustrating the use of a delay cell to cancel permanent echoes

required to give 1 msec delay. The delay cell is commonly employed not only to produce the delay necessary for permanent-echo cancellation but also to determine the pulse-recurrence frequency of the transmitted pulses. This is illustrated in the block diagram of Fig. 7.12.

F-m. Radar

In the circuit just described, range is determined in a Doppler radar by using amplitude modulation. An alternative method of determining range is to frequency-modulate the transmitted wave.

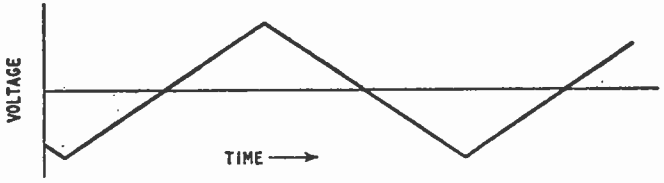


Fig. 7.13. Form of modulating wave employed in one type of f.m. radar

In this system the carrier wave is modulated by a sawtooth wave composed of alternate linear rises and falls as shown in Fig. 7.13. This causes the carrier frequency to vary linearly with time between two extreme values.

Due to the finite time taken by the wave to travel to the object and back, the wave picked up by the receiver has a frequency different from that being radiated by the transmitter at the instant of reception. This is illustrated in Fig. 7.14. The frequency difference Δf is linearly related to the distance of the object and if the difference frequency is indicated by a frequency meter, this instru-

ment can be calibrated directly in terms of the distance of the object.

The difference frequency can be obtained as a discrete component from the output of a mixer stage fed with a sample of the transmitter output and the received signal. The output amplitude of the signal received from the mixer depends on the amplitude of the received signal and this can vary over a wide range depending on the size and the distance of the object. Frequency meters require a

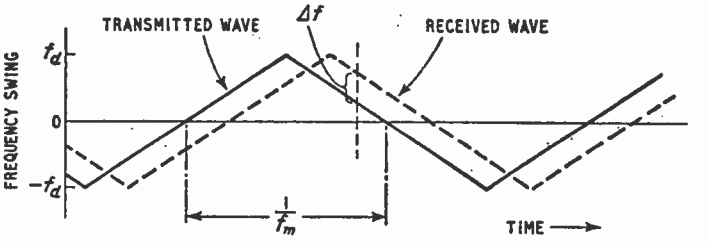


Fig. 7.14. Determination of frequency difference in an f.m. radar

constant-amplitude input to give accurate indications and thus the output of the mixer is fed first to an amplifier and then to a limiter stage before application to the frequency meter. A block schematic diagram of a range-finding radar of this type is given in Fig. 7.15. Equipments of this type are used in aircraft to indicate the distance of the ground below them: in this application the instrument is termed a radio altimeter.

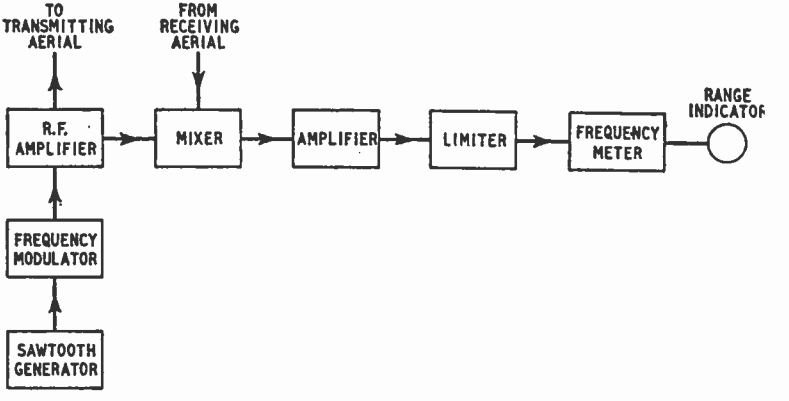


Fig. 7.15. Block schematic diagram for f.m. radar

The relationship between the distance of the object and the beat frequency Δf can be determined in the following way. Suppose the object distance is d. Then the radio waves will travel a total distance of $2d$ during the return journey to the object and back to the radar.

If the speed of radio waves is c, the time taken by the wave to travel this distance is $2d/c$ seconds. Suppose the modulation frequency is f_m and the frequency deviation (i.e. maximum frequency displacement) is f_d. Then the frequency of the transmitter changes by $2f_d$ in an interval equal to half the periodic time of one complete cycle of the modulating waveform. The frequency thus changes by $2f_d$ in $1/2f_m$ seconds. The rate of change of frequency is thus $4f_df_m$ c/s per second. In $2d/c$ seconds, therefore, the frequency changes by an amount Δf where

$$\Delta f = 4f_df_m \times \frac{2d}{C}$$

giving

$$d = \frac{c\Delta f}{8f_df_m}$$

Provided the deviation and modulation frequencies are constant, this frequency change is proportional to the object distance and thus the frequency meter can be directly calibrated in terms of object distance.

It is significant that the carrier frequency of the transmission does not enter into this expression and the radar accuracy is thus unaffected by changes in carrier frequency.

Detection of Faults in Cables and Waveguides

This application of frequency-modulated waves is also used to determine the position of faults in cables or waveguides. If there is any discontinuity in the characteristic impedance of a cable such as a short circuit or an open circuit, a wave sent along the cable is reflected at the point of the discontinuity, giving rise to a return wave in much the same way as an object reflects radio waves in open space. If, therefore, the forward-travelling wave is frequency-modulated by a sawtooth signal as described above, the reflected wave will give a beat frequency when heterodyned with the forward-travelling wave, and the value of the beat frequency is proportional to the distance between the sending end and the discontinuity. The formula given above can be used to calculate the distance of the fault. The expression includes c the velocity of propagation of

electromagnetic waves which, in this example, means the velocity in the cable or waveguide and this is usually appreciably less than the velocity in free space.

APPLICATION OF FREQUENCY MODULATION IN TELEGRAPHY

The principle of frequency modulation has extensive application in telegraphy, the science of sending intelligible messages over a distance. In general, we can distinguish two basic types of telegraphy; that in which signals are sent over a metallic link such as a cable to the far end of the circuit (this is line telegraphy), and that in which a radio link is used to carry the message (this is radio telegraphy).

In the earliest type of line telegraphic system the messages are sent by making and breaking a simple circuit such as that illustrated in Fig. 7.16. This may be regarded as an example of amplitude modulation in which the carrier frequency is zero. The current in the circuit alternates between two values, one of which is zero, enabling messages to be sent by means of the Morse code. Alternatively—and this is the system preferred today—the messages are

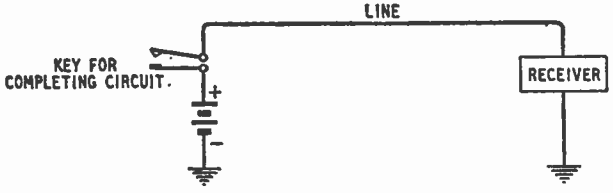


Fig. 7.16. Elements of a simple telegraphic system

sent out and received on teleprinter machines which print the messages automatically.

The machines are similar in appearance to ordinary typewriters and when a particular key is pressed, the appropriate character is printed on the paper (in capital letters only) and at the same time a message is sent along the line to the receiving end where a similar machine prints the same character. The machine is so constructed that each letter produces a characteristic set of impulses (each alternating between two levels only) which can be interpreted by the receiving machine.

In the simple amplitude-modulated telegraphic system just described only a single message can be sent over a given line at a

given time, but the system can be modified to accept a number of signals which can be sent simultaneously over the line without mutual interference. This is achieved by use of a number of carriers. Each telegraphic signal is used to amplitude-modulate a carrier, and the carrier frequencies are chosen so as to lie within the pass-band of the line but are so spaced that no mutual interference is likely to occur between neighbouring carriers. This system works well but is susceptible to interference. Any unwanted signals induced in the line by its proximity with other circuits are received as signal voltages which can cause spurious operation of the teleprinter. A great improvement in reliability is possible by frequency-modulating the carriers. The receiving equipment can now be made unresponsive to voltage fluctuations in the incoming signals, responding only to frequency fluctuations. A similar improvement in signal-to-noise ratio is obtained by adopting f.m. in place of a.m. in sound broadcasting.

In radio telegraphy messages can also be sent either by amplitude modulation of the carrier, alternating its amplitude between two values (one of which is often zero) or by frequency modulating it. The second method, often known as frequency-shift keying, can give the better signal-to-noise ratio.

APPLICATION OF FREQUENCY MODULATION IN FACSIMILE

Facsimile is that branch of telegraphy which deals with the transmission and reception of still pictures. This, too, may be achieved by means of a line or a radio link.

The principles employed in facsimile are similar to those used in television. The picture is assumed to be composed of a large number of elemental areas arranged in rows, each element having a definite tone which is uniform over its area. The term tone is here used in a photographic sense and means the degree of light or shade. The picture is scanned by an agent which traverses the whole area of the picture examining each element in turn and assessing its tone. The scanning agent produces an electrical output related to the tonal values and this output constitutes the signal which is transmitted to the receiving end of the circuit. Here the electrical signal is fed to a second scanning agent moving over a sheet of paper in exact synchronism with the movements of the first agent at the transmitter. The second agent is arranged to mark the paper so as to give tonal values suitably related to the electrical

input and in this way a reproduction of the original picture can be produced.

In one form of facsimile equipment the picture to be transmitted is wrapped around a drum which is rotated at a constant speed of about 100 revolutions per minute. At the same time the drum is moved along its axis by means of a nut which engages with a lead screw of 100 threads per inch. The scanning agent is a beam of light which is focused on the photograph so as to cover an area of approximately 1/100th inch diameter. The light reflected from this area of the photograph is collected by a photocell which gives an electrical output directly proportional to the light input and hence proportional to the tonal value of the element under the incident light at any moment. As the drum revolves the succession of varying voltages received from the photocell is used, after amplification, to frequency-modulate a carrier which can be transmitted to the receiving end of the circuit.

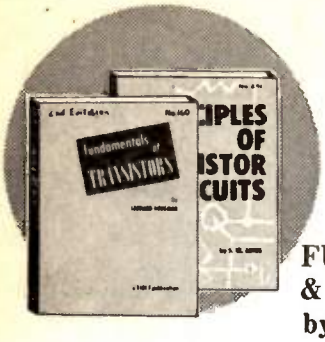
Here the output of the receiver is used to control the intensity of a lamp, light from which is focused on to an area of a piece of light-sensitive material such as photographic film or photographic paper wrapped around a drum which is rotated and driven axially in a manner similar to that at the transmitting end of the circuit. It is essential for successful results that the speed of the receiving drum should be precisely equal to that of the transmitting drum. Moreover, the two drums should also be in phase so that the scanning agents at both ends of the circuit cross the joining edges of the paper at the same instant.

This method of picture transmission is extensively employed by news agencies for the distribution of photographs for use in the press and here again the use of frequency modulation for radio transmission makes a great improvement in the signal-to-noise ratio and hence in the clarity of reproduction of the pictures.

INDEX

U. S. Price - $3.50

OTHER RIDER BOOKS



- *Accurate*
- *Dependable*
- *Easy-To-Read*
- *Profusely Illustrated*

FUNDAMENTALS OF NUCLEAR ENERGY & POWER REACTORS

by Henry Jacobowitz

After dealing with the fundamental concepts, the book turns to specific plants and discusses their construction, principles of operation, cost, and power output. Experimental reactors and the forerunners of the units now under construction are covered, as well as the current developments and future possibilities. Numerous diagrams and carefully selected photographs aid in the presentation.

CHAPTERS: The Atom and Its Nucleus . . . Nuclear Fission and the Chain Reaction . . . Basic Types of Nuclear Reactors . . . Power Reactors and Plants. Index.

Soft Cover, 6 x 9", illus., Catalog No. 218

FUNDAMENTALS OF TRANSISTORS (2nd Edition, Revised & Enlarged)

by Leonard M. Krugman, P.E.

This second edition (revised and expanded) modernizes the text so as to embrace the latest developments in the transistor art and so bring the book completely up-to-the-minute. The numerical examples contained in this book make every equation both understandable and usable.

CHAPTERS: Basic Semi-Conductor Physics . . . Transistors and Their Operation . . . The Grounded Base Transistor . . . The Grounded Emitter and Grounded Collector Transistors . . . Transistor Amplifiers . . . Transistor Oscillators . . . Transistors at High Frequencies. Appendix I: Patent References. Appendix II: Common Transistor Symbols. Index.

Soft Cover, 5½ x 8½", illus., Catalog No. 160

PRINCIPLES OF TRANSISTOR CIRCUITS

by S. W. Amos, B.S.C. (Hons.), Birmingham, England

This book serves as an introduction to the design of transistorized amplifiers, receivers, and numerous other electronic circuits.
Under the basic principles of transistors, the point-contact transistor and junction transistor are covered in very considerable detail. Items discussed include current amplification factor, voltage gain, power gain, alpha and alpha cutoff frequency.

CHAPTERS: Semi-Conduction and Junction Diodes . . . Basic Principles of Point-Contact and Junction Transistors . . . Common-Base Amplifiers . . . Common-Emitter Amplifiers . . . Common-Collector Amplifiers . . . Bias Stabilization . . . Small-Signal Amplifiers . . . Large-Signal Amplifiers . . . Transistor Superheterodyne Receivers . . . Other Applications of Transistors and other Types of Transistors. Index.

Soft Cover, 5½ x 8½", illus., Catalog No. 241

JOHN F. RIDER PUBLISHER, INC. 116 WEST 14th ST., N. Y. 11, N. Y.