MODERN RADIO TECHNIQUE

General Editor: J. A. RATCLIFFE

RECENT ADVANCES IN RADIO RECEIVERS

RECENT ADVANCES IN RADIO RECEIVERS

BY

L. A. MOXON

CAMBRIDGE
AT THE UNIVERSITY PRESS
1949

PUBLISHED BY
THE SYNDICS OF THE CAMBRIDGE UNIVERSITY PRESS

London Office: Bentley House, N.W. I
American Branch: New York

Agents for Canada, India, and Pakistan: Macmillan

Printed in Great Britain at the University Press, Cambridge
(Brooke Crutchley, University Printer)

PREFACE

The field covered by the title of this book is a wide one, and in such cases the selection of some of the material inevitably tends to be biased by the author's own personal experience and interests. It is hoped, nevertheless, to present the reader with a reasonably complete outline of recent developments and general trends.

The advances described have been due to the efforts of a large number of individuals working in Universities, Government Establishments and commercial firms. The credit for any particular achievement is often very difficult to apportion owing to the limited amount of published material so far available, and to the fact that frequently very similar results have been obtained and conclusions reached, more or less simultaneously and independently, by different investigators.

In supplementing my own knowledge and experience I have made considerable use of published material, mainly from British and American sources, and appropriate references are given in the text. I am also indebted to my colleagues and others, and particularly N. Houlding, for useful material and many helpful discussions.

Portions of this book have previously appeared in the *Journal of the Institution of Electrical Engineers*, the *Wireless World* (including the greater part of Chapter V) and *Nature* (figs. 2.1 and 2.2). Permission to make use of this material is gratefully acknowledged.

L. A. M.

CONTENTS

CHAPTER 1

Introduction

CHAPTER 2

The Concept of Noise Factor

CHAPTER 3

The Theory and Practice of Amplifier Design for Minimum Noise Factor

CHAPTER 4

The Noise Factor of Mixers

CHAPTER 5

The Measurement of Noise Factor

CHAPTER 6

Intermediate-Frequency Amplifiers

CHAPTER 7

Trends in Practical Receiver Design

CHAPTER 8

Some New Kinds of Receiver

CHAPTER 9

Some New Circuit Tricks

APPENDIX

Design Formulae for Intermediate Frequency Amplifiers

INDEX

Chapter 1

INTRODUCTION

The books in this series 'Modern Radio Technique' are intended for those who wish to bring their knowledge up-to-date, and it is assumed that the reader possesses at least a good working knowledge of Radio Technique as it was known in 1938 before the second World War. The more spectacular of the advances in the receiving field have been those associated with the opening up of the short-wave end of the spectrum, and the range of frequencies in common use has been extended upwards to hundreds of times its previous width. These wave-lengths have been for the most part devoted to new techniques such as radar and pulse-communication but this by no means exhausts their possibilities; for example, with the return of peace, one important new application is for the radio links needed in the establishment of nation-wide television services.

Receivers intended for any of these purposes usually have to meet exacting specifications in two respects; first, they must be capable of responding faithfully to impulses of very short duration, and secondly the internal noise level must be as low as possible, since this is the factor which normally sets the limit to the range of reception, whereas on the longer wave-lengths signal-to-noise ratio is more usually determined by atmospherics or other noise originating outside the receiver. This book is therefore concerned largely with these two major aspects of receiver design, but the last three chapters are devoted to other matters such as a review of general tendencies in the design of broadcasting, television, and communication receivers and a description of new kinds of receivers and circuitry. This portion covers a somewhat wider field than the earlier chapters, although in regard to these it will be appreciated that there is no watertight division between low-frequency and high-frequency techniques; for example, if it is necessary for some reason to use an inefficient aerial at medium wave-lengths (a situation which frequently has to be allowed for by designers of broadcast receivers) atmospheric noise ceases to be the limiting factor and the principles laid down in Chapter 3 may become just as important as they are at, say, 100 Mc./sec. and higher frequencies. The development of the shorter wave-lengths has of course stimulated intensive research into noise problems, leading to a much better understanding of the whole subject, and greatly improved performance of receivers; these advances are associated with the coming into general use, in 1941–2, of the noise-factor concept. With the problem studied from this angle, progress was greatly assisted, and the introduction of the grounded grid triode r.f. stage constituted an important step forward; the first use of this circuit for improving receiver noise factor was in British radar equipments operating in the regions of 600 and 200 Mc./sec., and it effected an improvement in the performance of the system equivalent to an increase in transmitter power of up to about four times. After this, attention was given to improvement of microwave radar systems, then at an early stage of development, and although nothing very spectacular was accomplished in the receiving field, steady progress in the manufacture and testing of mixer crystals, and in i.f. amplifier noise factor, eventually made possible an improvement of 4 db. or so in average performance.

The search for better noise factor has been accompanied by the evolution of greatly improved methods of sensitivity measurement, based on the use of noise sources, and this technique is now in process of extension to the microwave region.

Knowledge and experience of the design of i.f. amplifiers has been accumulated, with particular attention given to pulse response and to the achievement of the maximum product of stage gain and overall band-width in very wide-band amplifiers. Other interesting developments have included the taming of the super-regenerative receiver as described in another volume of this series.

In the following pages emphasis is placed on basic ideas rather than the solution of specific problems, and it has been assumed that the reader is aware of the basic principles of radar and television. An effort has been made to avoid unnecessary resort to mathematics and to convey ideas as far as possible by physical pictures. Formulae required for design purposes or to give precise expression to physical concepts are given, and their derivation explained except when it is readily available in other text-books or when the length and complexity of the required treatment is considered to be out of proportion to its value.

Chapter 2

THE CONCEPT OF NOISE FACTOR

2·1. The basic principles

The inherent noise level of a receiver, in conjunction with its aerial system, sets a lower limit to the strength of signal capable of being usefully received, unless other circumstances intervene, such as interference or insufficient amplification. The smallest useful signal, assuming adequate gain and no interference, depends on (*a*) the noise level, and (*b*) the complex technical and subjective considerations involved in the process of rendering signals intelligible to the senses.

These two sets of considerations can, as a first approximation, be treated independently, and our main concern for the moment is with noise level, although (*b*) requires further discussion in due course.

Noise originates partly in the aerial and partly in the receiver, and a receiver approximates to the ideal when its contribution is a negligible part of the total noise level. The number of times by which a receiver of adequate gain falls short of this ideal is the logical quantity to use as a measure of its sensitivity, and the term 'noise factor' has been introduced and defined on this basis. The average thermal noise voltage V_n across any resistance R at a temperature T is given by the well-known Nyquist formula

$$V_n^2 = 4kTBR, \tag{2·1}$$

where k is Boltzmann's constant, equal to $1{\cdot}380 \times 10^{-23}$ Joules/° K., and B is the band-width over which the noise is accepted. If the receiver noise can be represented by an equivalent noise voltage V'_n in the aerial, and if V_n is the aerial thermal noise, the noise factor clearly tends to unity as V'_n/V_n becomes small. So far the concept is an extremely simple one, but when we try to express it in the form of a *precise* and universally acceptable definition several difficulties arise. There are not only many alternatives to choose from, but also a number of pitfalls, the avoidance of which requires considerable caution, and the task is not made any easier by the historical background; the idea of noise factor occurred in the first instance to various workers more or less independently, quite a number of different definitions have been used, and there is still a certain amount of confusion. At the time of writing a precise and universally agreed definition does not exist, although reference to recent literature indicates considerable progress in this direction and divergences are not important in the majority of practical cases.

The concept of noise factor enables the sensitivity of receivers to be expressed on an 'absolute' basis and is valuable in computing the performance of various kinds of communication and radar systems, in facilitating comparisons of sensitivity, and in providing a sound approach to receiver design problems. It is subject, however, to certain limitations, and this, together with the other circumstances mentioned above, makes it necessary to discuss the various aspects of the definition of noise factor in some detail.

2·2. Source temperature

If the thermal noise of the aerial (or other source of signal) is to provide the reference level, it is important that the temperature of the source should be accurately known and stated. The 'temperature' of an aerial as judged by the noise voltage at its terminals is frequently an unknown quantity, and may vary in practice from only just above absolute zero, as may be found in the case of a microwave aerial with its beam directed upwards, to 20,000° K. or more for a typical metre-wave radar aerial looking at the noisiest part of the Milky Way. A 7 m. television aerial will be found to have an effective noise temperature of the order of 6000° K. due to galactic radiation, whereas at longer wavelengths unable to penetrate the ionosphere the noise temperature (excluding 'static') may not differ much from that of the earth's surface although very few data are available on this at the time of writing. In the case of receiver measurements in the laboratory the signal source is normally at room temperature, and this fact has led to the assumption of a standard temperature of 290° K. or thereabouts in some definitions of noise factor, variations of room temperature being unimportant for the degree of precision usually demanded.

In making accurate measurements it will be found, however, that the receiver noise and source noise are generally different functions of ambient temperature, and a further objection to specifying an arbitrary standard temperature arises if the receiver is intended to work from an aerial of known noise temperature substantially different from 290° K. It is probably advisable to abandon the idea of a standard temperature in favour of a stated temperature; for convenience this is usually in the region of

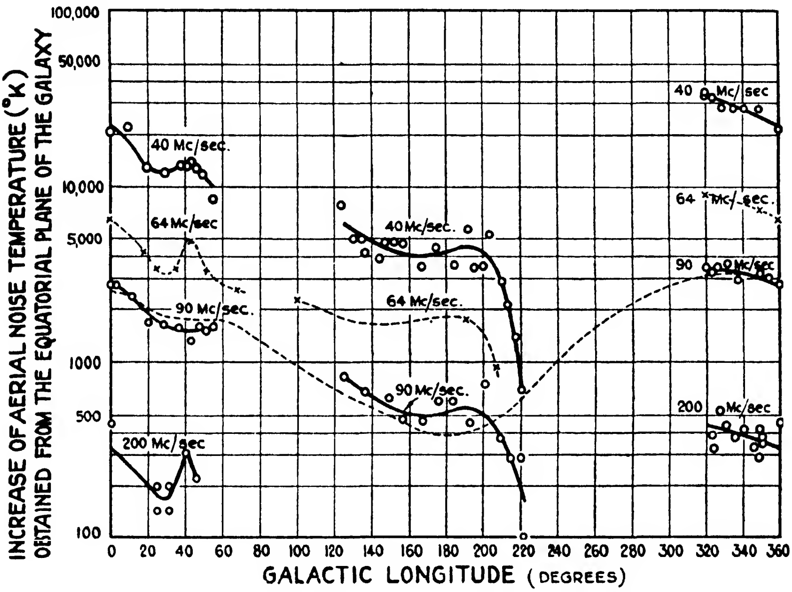


Fig. 2·1. Variation of aerial noise temperature with galactic longitude for a narrow beam aerial. The width of the noise peaks, to half temperature, is in the region of 15 to 30° of galactic latitude.

290° K., and in assessing the practical significance of noise-factor figures it is important to bear in mind that the actual temperature of the source may, as already stated, be very different. For example, noise radiation of extra-terrestrial origin (which was first observed by Jansky in 1932) has recently been subjected to detailed investigations covering a wide frequency range, and figs. 2·1 and 2·2 are based on various measurements of noise from the Milky Way in the frequency range of 40–200 Mc./sec.* This radiation is for the most part constant with time, and from

* For a detailed account of these measurements see *Nature*, 158 (1946), 759. The 64 Mc./sec. curve is based on the results of Hey, Phillips and Parsons, *ibid.*, 157 (1946), 296.

the curves it is possible to make a rough estimate of the noise temperature of the aerial if the beam-width and direction of the aerial and the relevant astronomical data are known. A receiver which generates, say, ten times the noise power fed into it from an aerial at 290° generates only 0·1 times the power fed into it at 7 m. from a narrow-beam aerial directed towards Sagittarius. Defining noise factor as a power ratio, and with respect to the appropriate source temperatures, we would obtain noise factors

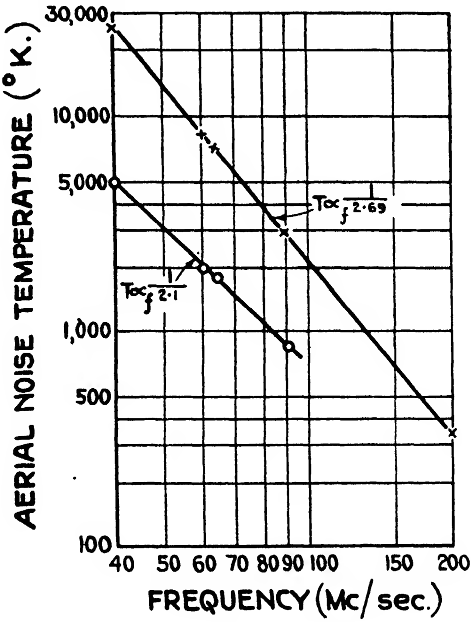


Fig. 2·2. Variation of galactic noise with frequency. The top curve corresponds to fig. 2·1, 340° galactic longitude, and the bottom curve is for aerials of, say, 20 to 60° beam-width directed away from the equatorial plane of the Galaxy.

of 11 times and 1·1 times respectively in the two cases; in the latter case the receiver is almost perfect, and in the former it falls far short of perfection.

Noise from the Galaxy should not, strictly speaking, be regarded as thermal noise, since it is most probably not of thermal origin, but it is very convenient to represent it by an 'equivalent aerial temperature', and its effect on the receiver is indistinguishable from that of thermal noise provided the band-width is narrow enough to justify the assumption of a constant mean noise level throughout the acceptance band. Noise from the sun sometimes reaches a much higher level than that from the Galaxy, but is ignored in this discussion because, owing to its irregular occurrence, it is hardly a relevant factor in receiver design.

2·3. Non-linearity in the receiver

It has been found by experiment that the useful sensitivity of a receiver does not depend appreciably on the law of the detector, whereas the apparent signal-to-noise ratio as measured on a meter obviously is so dependent. To be useful as a measure of performance noise factor must therefore be defined in such terms that it is independent of detector law. Alternative forms of definition all giving the same result include the following:

(*a*) The number of times by which the input-signal level required for a given output signal-to-noise ratio is increased as a result of the receiver noise, the output level being kept constant.

(*b*) The number of times by which the total output noise exceeds that portion of it due to the thermal noise of the source, the receiver having a linear input-output characteristic up to the peak output level of signal-plus-noise.

(*c*) The noise-to-signal ratio at the output divided by the signal-to-noise ratio at the input, reckoning the input noise over the 'energy band-width' (§ 2·8), the receiver being linear up to the peak output level of signal-plus-noise.

(*a*) has the advantage, so far as the definition is concerned, of avoiding the need for linearity in the receiver, and (*c*) has advantages in so far as the quantities involved are all readily measurable or calculable as such. The stipulation in regard to linearity sometimes takes the form of saying that, for the purpose of the definition, the output must be measured before detection, but this is unsatisfactory since amplifying stages in front of the detector may in practice be non-linear, and the definition should be based on measurement of output prior to any element in the receiver having a non-linear output-input characteristic. It must, of course, be measured *after* all important sources of noise. Usually it is more convenient to use a measuring device following the detector, and a simple experimental procedure is available which avoids the linearity difficulty, as described in Chapter 5.

2·4. The concept of available power

Noise factor has been variously defined as a ratio of voltages, powers and 'volts squared', but the first of these alternatives has not received much support. Probably the main objection to voltage ratio is due to the use of wave-guides, since current or voltage measurement in a guide presents difficulties whereas power is readily determined. Definitions in terms of available power, or square volts, are equivalent to each other except in circumstances to be discussed later, which normally arise only in conjunction with very wide band-widths.

Given a source of power having an open-circuit voltage E and internal impedance R, it is well known that maximum power is obtained from it when the load impedance is also R. The voltage across the load is then $\frac{1}{2}E$ and the power delivered is $E^2/4R$. The same result is obtained with a generator impedance $R + jX$ if the load impedance is $R - jX$. The quantity $E^2/4R$ is the power *available* from the generator, whether we use it by connecting a load of the right impedance or not. Applying the same procedure to the thermal noise generated in a resistance, the available noise power is $4kTBR/4R = kTB$, and is independent of the ohmic value of the resistance. This is a very important result, as it means that the ratio of available signal-to-noise power is not affected by any loss-free impedance transformations which it may be convenient to make in the aerial feeders or receiver input circuits. The use of available power, rather than power used, is desirable because best performance is usually obtained with an impedance mismatch between the source of signal and the receiver. It may be objected, perhaps, that since we are concerned only with a ratio the actual value of the load is unimportant, and it is true that a measurement of the ratio of powers used would in general give the right answer; on the other hand, it is the ratio of available powers which primarily decides the possibility of detecting weak signals, and the use of available power has the further merit that except when working with wave-guides the source impedance and open-circuit voltage are usually known, whereas the load impedance is not usually known to the same degree of accuracy.

With a receiver of noise factor N_r an available signal power in the aerial equal to $mN_r kTB$ would give a signal-to-noise power ratio of m as measured at any suitable point in the receiver, i.e. beyond the sources of noise but in front of any non-linear device. If a source of noise is used as the signal generator as described in Chapter 5, and if we define an equivalent noise temperature T' by putting $kT'B$ equal to the noise power available from the source in excess of the reference level kTB, we have $kT'B = mN_r kTB$, or in other words, $N_r = T'/mT$. By using the same 'B' throughout, we have, in effect, assumed that all the noise is amplified at the same band-width; this eliminates B and avoids the necessity for defining or measuring band-width. If it is required to measure noise factor by means of a c.w. signal generator, a determination of B (as defined in § 2·8) is essential.

2·5. Voltage squared versus available power

As already mentioned, it is possible to replace 'available power' by 'voltage squared'. This gives just the same result provided the source of signal possesses constant impedance throughout the band-width of the receiver, but the use of available power rather than voltage squared has so far been fairly general; this is because at the higher frequencies it is power rather than current or voltage which is usually measured, and also because it simplifies the treatment of impedance transformations which alter voltages but not the available power. Objections have been raised to the use of available power, on the grounds that it does not meet the case of variable impedance of the source, which is quite likely to arise, especially with very large band-widths. A source impedance of the form $R \pm jX$ requires the possibility of a matching impedance $R \mp jX$ if the available power concept is to have any meaning; unfortunately X, and usually R, are functions of frequency, and it is not possible to satisfy the required condition over a band of frequencies.

In defining the noise factor of a complete receiver it would be possible to stipulate constancy of source impedance, making allowance for any non-fulfilment of this condition when dealing with practical problems. Further difficulty arises, however, in considering partial noise factors (§ 2·10), and we are also faced with the common practice of defining noise factor with respect to a source 'having the impedance from which the receiver is designed to work'. The alternatives are to retain the available but also to any linear part of it. Such partial noise factors are useful, for example, in resolving the performance of mixer-amplifier combinations, and amplifiers of low stage gain. Noise factor may be expressed either as a numerical ratio or in decibels, sometimes one and sometimes the other being more convenient.

The wording of this definition has been carefully chosen to avoid having to couple it with a definition of energy band-width, but if this is known the 'additional noise power' of the definition may be replaced by a signal voltage in accordance with equation (2·1).

Noise factor as defined above is a satisfactory measure of the sensitivity of most types of receiver whose sensitivity is noise-limited, but there are exceptions, such as the crystal-video receiver (§ 8·3) and super-regenerative receivers operating in the non-linear or logarithmic mode.

2·7. Perception factor

As already mentioned, noise factor is not the whole story in regard to the effective sensitivity of a receiver, since the ability to detect weak signals depends not only on noise factor and aerial noise temperature but on a variety of complex, inter-related and largely subjective factors not amenable to precise calculation. By way of illustration it is useful to consider the particularly difficult case of a radar display system of the plan position indicator (P.P.I.) type, in which a radial trace rotates in synchronism with the aerial and signals are displayed as a brightness modulation; the detectability of weak signals is dependent on the size of the spot, the speed of the trace, the beam-width of the aerial and its speed of rotation, the width and shape of the transmitted pulses, the band-width and shape of the receiver response, the brightness of the picture, the illumination of the room and, not least, on the observer. The term 'perception factor' has been suggested to take account of all such variables, and although it has not yet come into general use it is necessary to define some such quantity in order to evaluate the overall performance of a receiving system. Thus, if D is the perception factor, T_a the temperature to which a resistance would have to be raised to make it produce the same amount of available noise power as the aerial, B the energy band-width as defined in § 2·8 in c./sec., and N_r the receiver noise factor relative to a source of signal at 290°, the minimum detectable signal in watts available from the aerial is given by

$$W_{\text{min.}} = kTB \times D \times \left(N + \frac{T_a - 290}{290}\right) = 4 \times 10^{-21} BD \left(N + \frac{T_a - 290}{290}\right) \tag{2·5}$$

D is in general dependent on B, and in the relatively simple case of a radar amplitude-time display the relation $W_{\text{min.}} \propto x + 2 + (1/x)$, approximately, where x is the product of band-width and pulse length, has been determined empirically.* If in this case the band-width is raised, for example, from the optimum value ($B = 1$/pulse length) to twice the optimum, thus doubling the noise power, we can see that the loss of sensitivity is only 0·5 db. If, on the other hand, the noise power is doubled at constant band-width, the loss is of course 3 db. The effect of any given component of the noise spectrum on the detectability of signals is therefore a function of its position relative to the acceptance band of the receiver, so that a unique description of the noise characteristics of the receiver (in terms of measurable quantities) requires knowledge not only of the noise factor but also of the noise spectrum. As a rule the noise spectrum coincides with the selectivity curve, but in the case of a receiver with low stage gains, several stages may contribute to the overall noise level and the band-width following the later of these stages will usually be wider than that following the earlier stages, so the resultant noise spectrum can be wider than the overall selectivity curve.

There are very few reliable data on the subject of perception factor, the necessary experimental work presenting considerable difficulty on account of the subjective elements which enter into it. It is possible, however, to make some rough generalizations in relation to r.f. or i.f. and audio or video band-widths, and detector law. Suppose, for example, that the transmission to be received contains events which take place in periods of time not shorter than T; obviously the information conveyed by these events will be lost unless the appropriate oscillations can be established in the circuits of the receiver during a time interval T. In other words the overall band-width must not be less than something of the order of $1/T$. On the other hand, if the band-width is made

* See, for example, A. V. Haeff, *Proc. Inst. Radio Engrs, N.Y.*, 34 (Nov. 1946), 857.

much wider than this, the total noise power increases proportionately, but very little more energy can be extracted from the signal and the overall sensitivity is decreased. From this line of reasoning it appears physically probable that the i.f. band-width required for maximum sensitivity will normally be of the order of $1/T$, and this seems to be borne out by such evidence as exists, including the example of radar reception* quoted above.

It might be expected that the same reasoning would apply to video band-width, but in fact this is not the case. If we again take the example of a radar amplitude-time display, we find that the video band-width is in no way critical from the point of view of sensitivity†; similarly, it is a matter of common observation that when receiving c.w. signals against a noise (as distinct from interference) background, narrowing of the band-width from, say, a kilocycle or so to perhaps two or three hundred cycles helps very little, if at all. In the latter case, at least, there is a simple physical explanation, because we all know that it is easy to pick out a weak sound from a much louder one *provided* the two are different in character. Noise at frequencies remote from that of a wanted beat note is therefore much less serious than the same amount of noise at the wanted frequency.

Why, then, does this argument not apply to i.f. band-width? The answer to this question is to be found in the mechanism of detection; each of the various components of the noise spectrum can be thought of as 'beating' with *all* the other components, so that when additional noise is brought in by widening the band-width it produces not only high frequencies but also additional low frequencies in the noise output of the detector.

The effect of detector law on the perception of weak signals in noise does not appear to be very significant. Both linear and square-law detectors are square-law in terms of signal-to-noise ratio at low ratios, because in the linear case the well-known phenomenon of modulation suppression replaces the exaggeration of amplitude differences by a square-law rectifier.

* It has been shown theoretically by Van Vleck and Middleton (*J. Appl. Phys.* Nov. 1946) that the optimum i.f. filter is the conjugate of the Fourier transform of the pulse, regardless of what happens after the detector.

† A. V. Haeff, loc. cit.

2·7·1. *Integration*

If sufficient time is available for making observations it is possible to obtain a large increase of sensitivity by the following methods.

Let us suppose, in the first instance, that the time available for observing the wanted signal is as short as possible, i.e. of the order of 1/(band-width). During this interval of time the instantaneous noise level may be anything from zero up to several times the mean level. Unless the level of signal-plus-noise is greater than the maximum probable instantaneous noise level, perhaps four times the mean, there is obviously no way of telling with sufficient certainty that a signal is present. On the other hand, the presence of a signal, even if the signal is very small compared with the mean noise level, will increase the average value of the receiver output level, and to detect the signal it is only necessary to make a sufficiently accurate measurement of the output level over a long period of time, and compare it with the mean 'no-signal' output level determined over a similar period. An application of this principle to the detection of very small increases in noise level is described in § 8·4.

Another application is for the reception of weak radar echoes. For this purpose it is necessary to 'gate' the receiver so that it is only operative during the time that an echo would arrive from a particular range and bearing. By integrating the output corresponding to a large number of pulses, the presence of a relatively weak signal can be detected, but a price, in the form of a slow rate of acquiring information, is paid for the increased sensitivity.

If the voltages due to n successive pulses can be added in like phase, before detection, the resulting amplitude is n times that of a simple pulse, whereas the noise being of random phase, only adds up to $\sqrt{n}$ times the amplitude. The improvement in signal-to-noise amplitude ratio would therefore be $\sqrt{n}$. It is very difficult, however, to carry out such addition before detection, and when allowance is made for the square law of the detector, which discriminates in favour of the noise when this is stronger than the signal, the improvement obtainable by integration after the detector is only $\sqrt[4]{n}$. Even so, an observation time of 10 sec. with a 1000-cycle recurrence rate gives an improvement of 20 db.

compared with observation of a single pulse. Some integration usually takes place, however, without any special measures, by virtue of the characteristics of cathode-ray tubes, since repetition of an echo several times in quick succession in the same place gives a brighter indication than can be obtained from a single trace.

2·8. Energy band-width

A knowledge of the energy band-width of a receiver is required for various purposes including assessment of overall performance (equation (2·5)) and measurement of noise factor by the c.w. signal generator method (equation (5·1)) which has not yet been superseded in all cases by the noise generator methods described in Chapter 5.

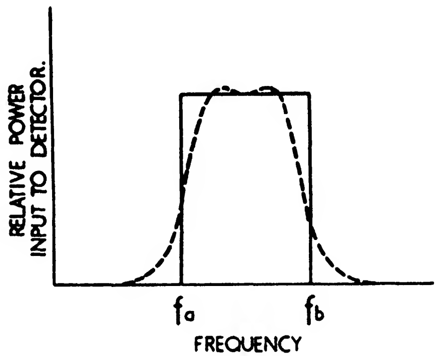


Fig. 2·3. Illustration of the concept of energy band-width. The areas under the two curves are equal and the energy band-width is $f_b - f_a$ in each case.

Fig. 2·3 shows a so-called 'ideal' rectangular selectivity curve. A receiver with such a response will amplify at maximum gain all noise energy reaching it between the frequency limits f_a and f_b and will be completely dead to all other noise. $f_b - f_a$ is therefore the band-width of the receiver from the point of view of noise energy.

A more typical curve for a practical receiver is shown dotted on the same scale, and it will be noticed that, assuming a uniform noise spectrum, less noise energy will be received between the limits f_a and f_b, but the deficiency will be made up by the response to frequencies outside these limits. If the area under the two curves is the same, then the energy band-width B is the same. The procedure for determining B is therefore to draw the selectivity curve in the form of relative power input P to the detector versus frequency, find the area under it (either with a planimeter or by counting squares) and divide by the height at the central (or carrier) frequency f_0. The quotient is the width of the equivalent rectangular response, or in other words the energy band-width.

In mathematical language, if P_0 is the height of the dotted curve in fig. 2·3 at f_0, the energy band-width is given by

$$B=\frac{1}{P_0}\int_0^\infty P df \qquad (2\cdot 6)$$

2·9. Input matching

A theoretical study of the design of receivers for best noise factor will be found in the next chapter. It is shown that as the impedance of the source is varied relative to the receiver input impedance, there is a condition for minimum noise factor which in general differs from that for maximum power transfer, and gives a noise factor which can approximate to unity. We are not interested here in gain, since a decibel or so more or less gain in a total of perhaps 100 db. is well within the normal tolerances to be expected as a result of differences in valves, and even if an additional stage of amplification were to be needed it would not as a rule present much difficulty. In the case of a receiver intended only for strong signals the situation is quite different, of course, and the input circuit design is usually based on considerations of selectivity or gain.

In the absence of a full appreciation of the implications of noise-factor theory it is natural to assume that the input circuit should be designed for maximum gain, and noise factor has sometimes been defined accordingly. Such definitions assume not only a 'power match', but also that the receiver input impedance is at the same noise temperature as the source; it must therefore contribute just as much noise as the source, and the nearest approach to perfection which can be achieved in this case, in terms of the previous definition, is a noise factor of 2. The convention whereby such a receiver is said to have unity noise factor has now been abandoned by common consent, but it is not unusual to find it stated that receiver noise factor 'cannot be better than 3 db.'; this limitation is only applicable to the special circumstances just described, and noise factors better than 3 db. are frequently achieved in practice.

2·10. Partial noise factors

Consider the case of a number of four-terminal networks in cascade as illustrated in fig. 2·4. Available power gains G_1, G_2, etc., respectively and noise factors N_1, N_2, etc., can be assigned to these networks in accordance with the definitions given in §§ 2·4 and 2·6. It will be assumed that all noise contributions are integrated over the same band-width B, and that all the noise factors are based on the same stated temperature T.

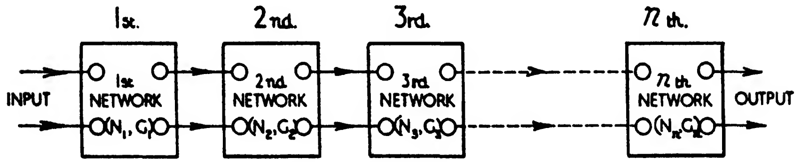


Fig. 2·4. Illustrating connexion of n networks in cascade.

If G is the overall gain of the chain of networks, the total noise power available at the output terminals is made up as follows:

Input noise power from source $= kTBG$

Additional noise contributed by first network $= kTBG(N_1 - 1)$

Additional noise contributed by second network $=$

$$kTBG(N_2 - 1)\frac{1}{G_1}$$

Additional noise contributed by third network $=$

$$kTBG(N_3 - 1)\frac{1}{G_1 G_2}$$

Additional noise contributed by nth network $=$

$$kTBG(N_n - 1)\frac{1}{G_1 G_2 \ldots G_n}$$

By definition of noise factor, these various contributions must add up to $kTBGN_r$, where N_r is the overall noise factor, and dividing both sides of the equation so obtained by $kTBG$ we obtain the result given by Friis:*

$$N_r = N_1 + \frac{N_2 - 1}{G_1} + \frac{N_3 - 1}{G_1 G_2} + \ldots + \frac{N_n - 1}{G_1 G_2 \ldots G_n}. \qquad (2·7)$$

* Loc. cit.

Even with the widest band amplifiers at present in use, it is unusual for more than the first three stages to require consideration. As a rule the assumption that the same band-width B is applicable for all the noise contributions is justifiable, but as already explained it may not hold for special cases, or very wide band-widths, and if the noise produced in networks 1, 2, 3, etc., is integrated over different energy band-widths B_1, B_2, B_3, etc., the contributions must be weighted accordingly. In trying to express this statement in precise terms there arises the difficulty of definition discussed in § 2·5. If, however, N_1, N_2, etc., are based on source impedances constant over the acceptance bands B_1, B_2, etc., the weighting factors are B_1/B, B_2/B, etc., respectively, and equation (2·6) becomes

$$N_r = 1 + (N_1 - 1)\frac{B_1}{B} + \frac{N_2 - 1}{G_1}\frac{B_2}{B} + \frac{N_3 - 1}{G_1 G_2}\frac{B_3}{B} + \ldots, \text{ etc.} \qquad (2{\cdot}8)$$

Chapter 3

THE THEORY AND PRACTICE OF AMPLIFIER DESIGN FOR MINIMUM NOISE FACTOR

3·1. Introduction

From what has been said in the previous chapter, it will be appreciated that noise factor is a property of *any* four-terminal network having a linear input-output characteristic, and is important when the network is, or is followed by, an amplifier of sufficiently high gain. The two cases in which we are most interested are (*a*) high-gain receivers with r.f. amplification, and (*b*) mixers followed by high-gain i.f. amplifiers. The mixer and amplifier problems are somewhat different in regard to the treatment they require; it is therefore proposed to separate them and deal in this chapter with amplifiers and in the next with mixers. When an amplifier is preceded by a mixer, the overall noise factor may be obtained from the separate mixer and amplifier noise factors by application of equation (2·7) using the first two terms.

3·2. General principles

In any receiver there is a certain amount of noise contributed by shot effect in valves, thermal noise due to the finite resistance of coils, etc., and noise associated with electron transit time.* Approximation to unity noise factor requires that these noise contributions shall be negligible compared with the thermal noise of the source, and this in turn requires the following physically obvious conditions to be satisfied:

(1) The gain of the first stage should be high enough to ensure that any noise originating later in the receiver is not amplified sufficiently to become appreciable.

(2) The thermal noise voltage of the source, as applied to the first valve, must be sufficient to modulate the electron stream

* At low frequencies there are other sources of noise, e.g. 'flicker effect', but such noise is of different character and the concept of noise factor is therefore difficult to apply.

by an amount large compared with the shot noise fluctuations. It is usual to specify valve noise in terms of the 'equivalent noise resistance' R_n, which is defined arbitrarily in accordance with equation (2·1) by equating $\sqrt{(4kTBR_n)}$, where T is 290° absolute to the noise voltage which, applied between grid and cathode, would produce an effect equivalent to the shot noise; then if R_a is the impedance of the source as seen by the valve, its thermal noise voltage [given by $\sqrt{(4kT_aBR_a)}$] is required to be large compared with $\sqrt{(4kTBR_n)}$, i.e. we require $R_a \gg R_n$.

(3) The fulfilment of condition (2) requires in general the use of some sort of impedance-transforming network for coupling the source of the valve. Any losses in this network will not only be generators of thermal noise but also absorb power from the source and thereby reduce the aerial thermal noise voltage at the grid.

(4) Electron transit time must be short, since it results in an input shunt resistance possessing a relatively high noise temperature.

At frequencies up to a few megacycles with narrow band-widths, provided the issue is not confused by such considerations as selectivity or the covering of a wide tuning range, it is possible to achieve noise factors negligibly greater than unity. As the frequency becomes higher the circuit losses tend to increase more or less directly with, and the loss due to transit time as the square of, the frequency. Feed-back effects also become more pronounced and can be harmful. The presence of input circuit or transit-time losses leads, as shown in the following sections, to a definite optimum value for R_a and a corresponding minimum value of noise factor.

In very wide-band amplifiers, noise factor is relatively high for several reasons. It is found, for instance, that the optimum value of R_a in conjunction with the valve-input capacity results in a much narrower band-width than is required, and R_a has to be reduced accordingly; the stage gain is low, and noise from the second and even later stages may be appreciable in consequence; and it is often necessary to use a relatively high mean frequency for reasons such as the one discussed in § 6·4·3.

3·3. Equivalent circuit for first stage with no feed-back

The general ideas outlined above may be given quantitative expression by devising equivalent circuits which depict the correct relationship between the various voltages and impedances, and subjecting them to mathematical treatment. This has been the basis of a number of more or less independent theoretical investigations by various workers, including published papers by D. O. North* and by E. W. Herold and L. Malters.† Use is made of equivalent networks in which the resistances or conductances can be regarded as noiseless, and the sources of noise and signal are represented by separate voltage or current generators. These treatments are necessarily of an approximate nature and make the assumption that there is no coherence between the various sources of noise. This is not strictly true, as the valve-shot noise is in fact coherent with the noise associated with transit time losses, but as first pointed out by North, with a resistive input circuit and small transit angles the relation is one of phase quadrature and the assumption is allowable. On the other hand, it might be expected that some small degree of noise cancellation could be obtained by de-tuning the input circuit to take advantage of the phase relationship, and this has been confirmed by experimental observations.

An accurate treatment can be based on the fundamental electronic equations of Llewellyn and Peterson,‡ but is complicated and its value is limited by practical difficulties in application. Provided the limitations are borne in mind, more value is likely to be obtained from the simpler approach to the subject in most instances.

In fig. 3·1 the source is assumed to be coupled to the valve through a transformer designed so that the valve looks into a resistive impedance R_a, all reactive components of impedance being eliminated by tuning. V_{si} represents the open-circuit r.m.s. signal voltage and $\sqrt{(4kT_aBR_a)}$ the open-circuit r.m.s. noise voltage of the source after transformation and, as explained in Chapter 2, the signal power and thermal noise power available

* *R.C.A. Rev.* 6 (1942), 332.

† Ibid. p. 302; also *Proc. Inst. Radio Engrs, N.Y.*, 31 (1943), 491.

‡ *Proc. Inst. Radio Engrs, N.Y.*, 32 (1944), 144.

from the source are respectively $V_{si}^2/4R_a$ and kT_aB, and are independent of the value of R_a. Circuit losses are accounted for by the shunt resistance R_c at temperature T_c, which may be different from T_a, and the associated noise is represented by the separate generator of e.m.f. equal to $\sqrt{(4kT_cBR_c)}$. The generator of voltage $\sqrt{(4kTBR_n)}$ represents shot noise as explained in § 3·2. It is assumed that the band-width of the input circuit is sufficiently wide for reactance effects to be negligible throughout the overall acceptance band B_0.

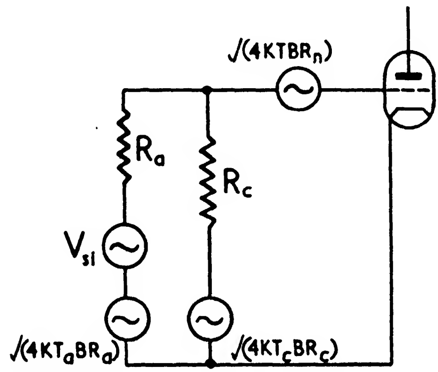


Fig. 3·1. Equivalent input circuit illustrating the effect of valve noise and circuit losses.

Representation of the effect of transit time by including in R_c the resistive component of valve impedance to which it gives rise, with a suitable 'temperature' assigned to R_c to account for the additional noise, is described in § 3·5.

None of the generators shown in fig. 3·1 has any internal impedance of course, and it is important to note that the 'equivalent noise resistance' of the valve is *not* a resistance, but a mathematical device which is convenient because of the simplification which results from having all the noise sources described in the same language.

When feed-back is present the equivalent circuit requires to be elaborated as explained in § 3·6, but provided the feed-back is such as to affect the signal and all sources of noise by the same amount the analysis which follows, based on fig. 3·1, remains applicable. This analysis only applies of course to a single stage, and if there is an appreciable noise contribution from several stages each must be treated separately and the results combined by means of equation (2·6).

3·4. Analysis of equivalent circuit with no feed-back

Referring to fig. 3·1, the signal voltage V_{so} applied to the valve is equal to $V_{si}R_c/(R_a+R_c)$. The noise voltage at the same point is made up of $\sqrt{(4kTBR_n)}$, $R_a\sqrt{(4kT_cBR_c)}/(R_a+R_c)$ and $R_c\sqrt{(4kT_aBR_a)}/(R_a+R_c)$, the total mean-square noise voltage V_{no}^2 being the sum of the squares of these three components since the relation between them is assumed to be random. The input noise-to-signal voltage ratio squared is $4kT_aBR_a/V_{si}^2$. Applying this data to equation (2·2), we obtain the result:

$$N_r = 1 + \frac{T_cR_a}{T_aR_c} + \frac{TR_n}{T_aR_a}\left(1 + \frac{R_a}{R_c}\right)^2 \tag{3·1}$$

by simple algebra.

An alternative method of deriving this expression follows directly from the definition of noise factor given in § 2·6. Thus, if an additional noise power equal to N_rkT_aB is made available from the source, the corresponding extra voltage squared applied to the valve is $4kT_aBR_aN_rR_c^2/(R_a+R_c)^2$, and by definition this must be equal to V_{no}^2. Writing out this equation in full and dividing both sides by

$$4kT_aBR_aR_c^2/(R_a+R_c)^2,$$

we obtain equation (3·1), proving incidentally that subject to the assumption of constant and purely resistive impedances throughout the pass band the two alternative forms of the definitions are identical.

The above method of treatment has been criticized on the ground that the use of conductances and current generators in place of resistances and voltage generators would make for tidier algebra, but it is probable that most of us are more used to thinking in terms of resistances than conductances, and, since the object here is to try and convey the clearest possible physical picture, use of the resistance and voltage convention is felt to be justified.

As explained in § 2·2, the definition of noise factor usually assumes $T_a = T$. Putting this condition in equation (3·1), we can see that minimum noise factor requires R_n and T_c/T to be as small and R_c as large as possible, subject to the inherent limitations of valves and circuits. We are left with a measure of freedom in the choice of R_a and find that as this is varied N_r is a minimum when

$$R_a = R_c \sqrt{\frac{R_n}{SR_c + R_n}}, \tag{3·2}$$

where $S = T_c/T$. This result does not depend on the temperature of the source. Substituting the value of R_a given by equation (3·2) and the conventions $T_c/T = S$ and $T_a = T$ in equation (3·1), we find that the minimum value of noise factor is

$$N_{r(\min.)} = 1 + 2\frac{R_n}{R_c} + 2\sqrt{\left[\frac{R_n}{R_c}\left(S + \frac{R_n}{R_c}\right)\right]}. \tag{3·3}$$

The band-width to 3 db. down of a tuned circuit having a capacity C and damped by a parallel resistance R is given by the well-known expression $1/2\pi CR$. In the case of the input circuit represented by fig. 3·1 the band-width is determined by R_a and R_c in parallel; in other words (assuming a simple tuned circuit of capacity C), it is given by $(R_a + R_c)/(2\pi C R_a R_c)$. If R_a is chosen according to equation (3·2) it may well be found that the input circuit band-width is narrower than the overall band-width required for proper reproduction of the signal. This point is further discussed in § 3·8.

3·5. Input loss and temperature of input impedance

Up to now we have assumed that R_c represents circuit loss alone. In an actual case it also includes losses due to the finite transit time of the electrons in the valve and may be conveniently separated into two components, R_t due to electron transit time, and R_e in parallel with R_t and including all other losses. We then have $R_c = R_e R_t/(R_e + R_t)$.

The simplest and most usual form of input transformer makes use of a single-tuned circuit, such as one or other of the arrangements shown in fig. 3·2. The value of R_a in terms of the actual source impedance R_s and of the circuit reactances is given in table 1, and R_e may be determined experimentally by measuring the circuit impedance with the valve connected to it but cold, R_s being removed and resonance maintained (if necessary) by connecting a suitable reactance (or in case (*a*) a short circuit) in place of it. Alternatively, if the distribution and magnitude of losses is known, R_e can be calculated. As a rough approximation R_e is usually given by $Q/\omega C$ where Q is the reactance of L divided by its internal resistance. The value of Q usually achieved in

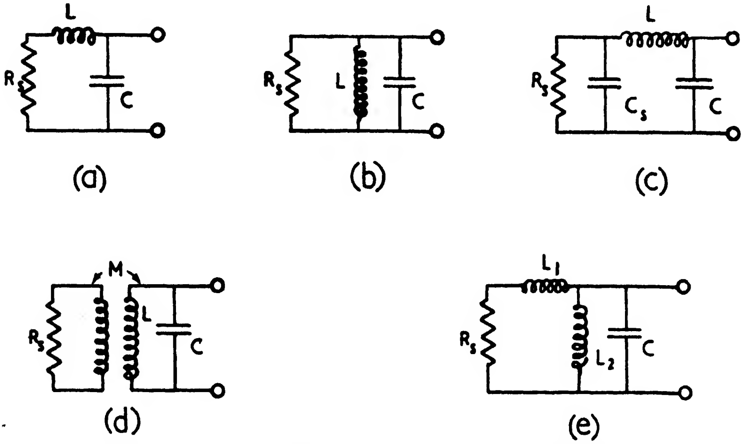


Fig. 3·2. Some alternative forms of input transformer.

practice is about 100 and does not vary much with frequency up to a few hundred megacycles. This means that R_e is roughly proportional to wave-length.

Table 1

Fig. 3·2	(*a*)	(*b*)	(*c*)	(*d*)	(*e*)
R_a	ω^2L^2/R_s	R_s	$\frac{C_s^2}{C^2}(R_s+1/\omega^2C^2R_s^2)$	L^2R_s/M^2	$\omega^2L_1^2/R_s$

Transit time loss arises as follows. As the electrons approach the grid they induce a charge on it, and the corresponding current is 90° out of phase with the grid voltage. When the transit time is very small this is cancelled by an equal and opposite current due to the electrons leaving the grid. If, on the other hand, the time taken by the electrons to travel from cathode to anode is comparable with the reciprocal of the frequency, we find that although the grid potential determines the number of electrons *leaving the cathode* at any instant, the density at a distance from the cathode depends on the potential which existed at an earlier instant, and in general the number of electrons approaching the grid will differ from the number leaving it so that there is a resultant grid current in quadrature with the grid voltage. We find also that the difference between the total number of electrons approaching and the number leaving the grid at any instant does not exactly follow the grid potential but is subject to a phase shift which shows

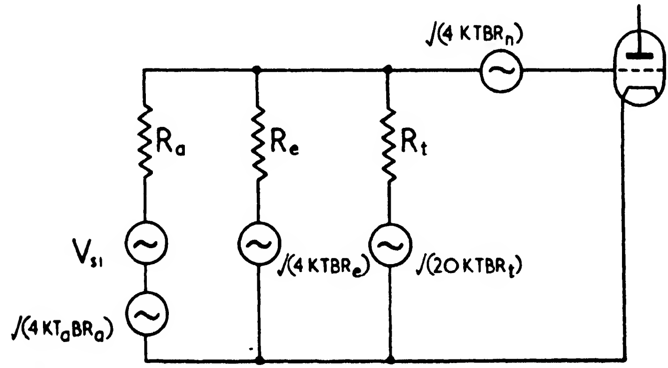


Fig. 3·3. Equivalent circuit showing subdivision of input losses into cold loss R_e, and transit time loss, R_t.

itself as a resistive or 'lossy' component of the grid-input admittance, and since both the magnitude of the grid current and its phase angle are proportional to frequency (for a given transit time) the loss increases as the square of the frequency. The random fluctuations of the induced current are in phase quadrature with the equivalent shot-noise voltage so long as the transit angle is small, and therefore the voltage they develop across a resistive input circuit can be combined with the shot noise by ordinary r.m.s. addition. North and Ferris* have shown that the equivalent temperature to be assigned to R_t to account for this noise is approximately five times room temperature for valves with oxide-coated cathodes. R_t decreases as the square of the frequency,† so that as the wave-length is reduced it becomes more and more important relative to R_e (which decreases directly as the frequency if Q is constant).

Fig. 3·3 shows R_e and R_t with their associated noise generators.

* *Proc. Inst. Radio Engrs, N.Y.*, 29 (1941), 49.
† W. R. Ferris, Ibid. 24 (1936), 82.

For the purpose of applying equations (3·1) or (3·3) we can replace R_e and R_t by the single resistance R_c, with an effective noise temperature derived as follows, assuming R_e to be at room temperature T. If R_a is infinite, the noise voltages squared at the input to the valve due to R_e and R_t are

$$4kTBR_e\left(\frac{R_t}{R_e+R_t}\right)^2 \quad \text{and} \quad 20kTBR_t\left(\frac{R_e}{R_e+R_t}\right)^2,$$

respectively. Replacing R_e and R_t by R_c, the voltage squared would be $4SkTBR_c$, and equating this to the sum of the voltages produced by R_e and R_t,

$$4SR_c=4R_e\left(\frac{R_t}{R_e+R_t}\right)^2+20R_t\left(\frac{R_e}{R_e+R_t}\right)^2.$$

Remembering that $\quad R_eR_t/(R_e+R_t)=R_c,$

this gives $\quad S$ or $(T_c/T)=(R_t+5R_e)/(R_t+R_e). \qquad (3·4)$

It has been assumed for simplicity that R_a is infinite, but when this is not the case both noise voltages are affected equally and the result is the same.

With this substitution, equations (3·1) to (3·3) provide useful guidance in the design of amplifiers. The main defects of this treatment are failure to predict what happens when the input circuit is not resistive, particularly the improvement which may be obtainable by antiphasing the shot noise and transit-time noise, and its failure to hold for large transit angles. The first of these effects appears to be small, and the second is of limited practical importance because when the breakdown point is reached the noise factor is already poor and better performance can be obtained by using a crystal mixer with amplification at i.f. only.

3·6. The effect of feed-back

Feed-back of one sort or another becomes increasingly difficult to avoid as the frequency is increased. If noise, whatever its origin, and signal are all fed back in equal amounts, noise factor is obviously not directly affected although it will be influenced indirectly, if there is appreciable noise from later stages, by the change of gain. An important instance of this kind of feed-back is provided by the grounded-grid triode circuit described in § 3·7.

If the signal and only some of the noise is fed back in antiphase, the result is an increase in the relative importance of the noise which has not been fed back, and the noise factor is made worse; conversely, positive feed-back of the signal and only part of the noise improves the noise factor, but this is of limited practical importance because of the engineering difficulties in maintaining a sufficiently high and yet stable level of positive feed-back. A particularly important instance of partial negative feed-back at high frequencies occurs in the ordinary r.f. pentode amplifier, and the following treatment of it is an extension of the methods used in §§ 3·4 and 3·5.

3·6·1. *The radio-frequency pentode*

It is well known that the input conductance of receiving valves of normal construction arises largely from the effect of inverse feed-back via the cathode-lead inductance. In the case of a triode this reduces the signal and the cathode noise current by the same amount and therefore does not affect the noise factor. With a pentode, however, additional anode current fluctuations are produced by the random partition of the electron stream between anode and screen. Since the cathode lead carries the total valve current, it is evident that the feed-back takes no cognizance of the division of current between anode and screen; the cathode noise and the partition noise should therefore be represented by separate generators corresponding to noise resistances R_{nc} and R_{ns} for the cathode and screen respectively, and the effect of feed-back is faithfully represented by a resistance R_f whose function is to reduce the signal voltage, etc., relative to the partition noise, but which is not itself a source of noise. This argument leads to the equivalent circuit of fig. 3·4. Analysis of this, though laborious, involves only simple algebra, and it will be sufficient to state the result which reduces to the previous equations (3·1) and (3·3) on applying the appropriate assumptions.

It is convenient to retain the symbol R_n for the total noise resistance $R_{ns}+R_{nc}$, and to replace R_e and R_t by R_c, with the noise temperature ratio S assigned to it in accordance with equation (3·4). Then

$$N_r = 1+\frac{2R_n}{R_c}+\frac{2R_{ns}}{R_f}+\frac{R_a}{R_c^2}\left\{SR_c+R_n+\frac{2R_{ns}R_c}{R_f}+\frac{R_{ns}R_c^2}{R_f^2}\right\}+\frac{R_n}{R_a}. \tag{3·5}$$

This is of the form $N_r = a + bR_a + c/R_a$, and as R_a varies N_r will be a minimum when $bR_a = c/R_a$. Therefore

$$R_{a\,(\mathrm{optimum})} = R_c \sqrt{\frac{R_n}{SR_c + R_n + R_{ns}\,(2R_c/R_f + R_c^2/R_f^2)}}, \tag{3·6}$$

and substituting (3·6) in (3·5),

$$N_{r(\mathrm{min.})} = 1 + \frac{2R_n}{R_c} + \frac{2R_{ns}}{R_f} + 2\sqrt{\left[\frac{R_n}{R_c}\left(S + \frac{R_n}{R_c} + \frac{2R_{ns}}{R_f} + \frac{R_{ns}R_c}{R_f^2}\right)\right]}. \tag{3·7}$$

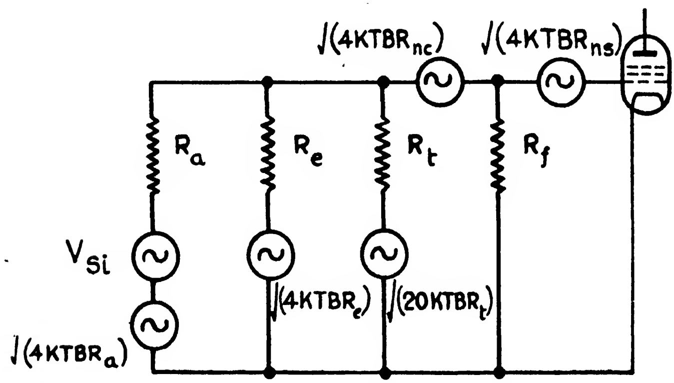


Fig. 3·4. Equivalent circuit of pentode showing effect of cathode-lead feed-back.

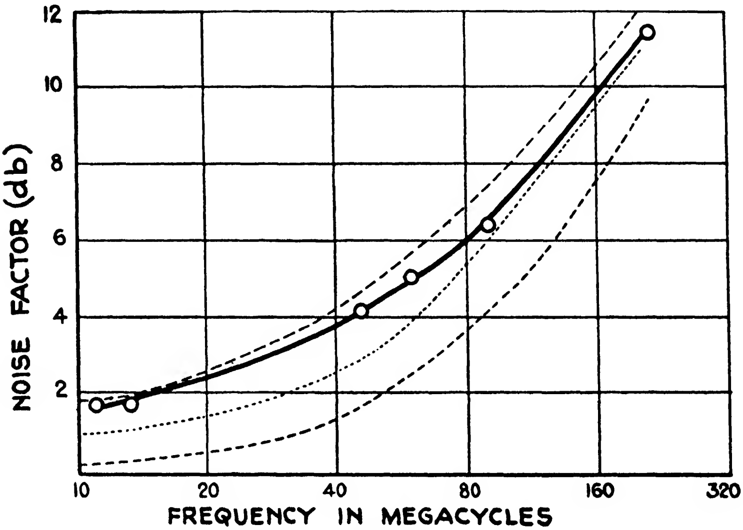
Serious difficulty arises in trying to apply equation (3·7) because figures available for valve-input impedance are the resultant of R_f and R_t. Both these resistances are inversely proportional to the square of the frequency, and it is very difficult to separate them experimentally. With available data, only a qualitative treatment of the problem is possible, and as a first step it is interesting to investigate the two limiting conditions obtained respectively by letting the valve-input impedance be due (*a*) entirely to cathode feed-back ($S = 0$) and (*b*) entirely to transit time ($S = 5$), with negligible coil losses in each case. For these two conditions the minimum value of N_r is given respectively by

$$N_r = 1 + \frac{2R_{ns}}{R_f} + 2\sqrt{\frac{R_n R_{ns}}{R_f^2}} \tag{3·8}$$

and

$$N_r = 1 + \frac{2R_n}{R_t} + 2\sqrt{\left[\frac{R_n}{R}\left(5 + \frac{R_n}{R_t}\right)\right]}. \tag{3·9}$$

Taking values of R_n and R_{ns} from table 2, equations (3·8) and (3·9) are plotted in fig. 3·5 for the RL7 pentode, assuming the input resistance to be $10^7/f^2$ ohms, where f is the frequency in megacycles/second. The right result must lie between these two



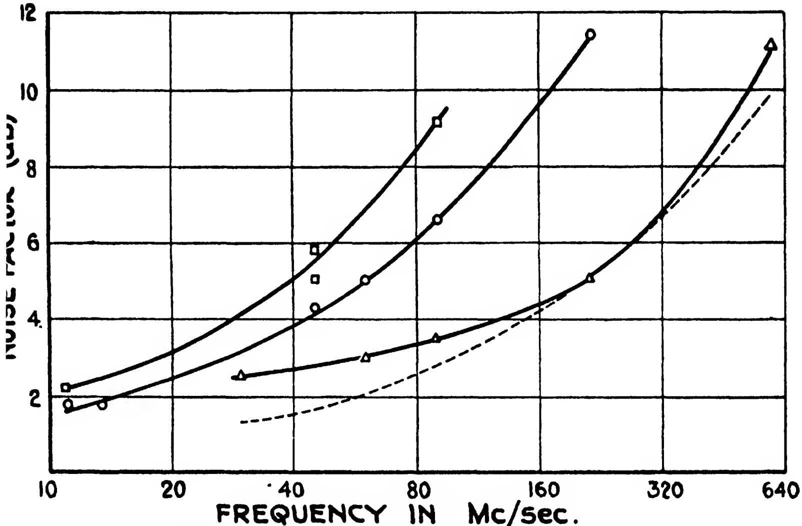
— experimental curve.
= boundary conditions from equations (3·8) and (3·9).
= theoretical curve from equation (3·1) putting $T_a = T_c = T$.

Fig. 3·5. Noise factor of RL 7.

limiting conditions unless coil losses are appreciable, as will tend to be the case at the lowest frequencies. A curve showing experimental results obtained with typical circuits is included in fig. 3·5, and these tend to be somewhat higher than the average of the values given by equations (3·8) and (3·9) especially at low frequencies, but to no greater extent than could be accounted for by neglect of coil losses.

It is evident from fig. 3·5 that N_r is not very critically dependent on the ratio R_f/R_t, and a rough estimate of this should therefore be adequate for most purposes.

Of the quantities which enter into equation (3·5), R_n and R_{ns} are to a certain extent under the control of the valve designer, and progress made in reducing them is described in § 3·10. R_f, being due to the inductance of the cathode lead, is also primarily a valve-design problem and can be almost eliminated at the expense of mechanical convenience by using the technique of disk seal construction; with conventional constructions it can be minimized by suitable valve design as described in § 3·10, but is the concern



= EF 50.
= RL 7 (EF 54).
= CV 53/S 25 A.
= noise factor of CV 53 predicted by fitting the universal curve shown in fig. 3·16 to the measured curve at 214 Mc./sec.

Fig. 3·6. Comparison of measured performance of EF 50, RL 7, and CV 53.

of the circuit designer as well, since it depends on the way in which connexions are made to the valve. R_c could usually be increased considerably by adopting much more bulky and expensive coil designs, but the improvement of noise factor would in general be relatively small. S depends primarily on the cathode temperature, and very little can be done towards reducing it. This leaves R_a as the most important variable of those under the control of the circuit designer; an optimum value can be readily found in accordance with equation (3·6) and is obtainable by proper design of the input transformer *unless* band-width considerations intervene as already explained (§ 3·4).

3·6·2. *Use of feed-back to improve the noise factor of pentodes*

It has been shown that the effect of negative feed-back via the cathode lead of pentodes is to make the noise factor worse. Conversely, it might be expected that positive feed-back could be used to improve the noise factor. This possibility has been considered qualitatively with the simplified circuit of fig. 3·7.

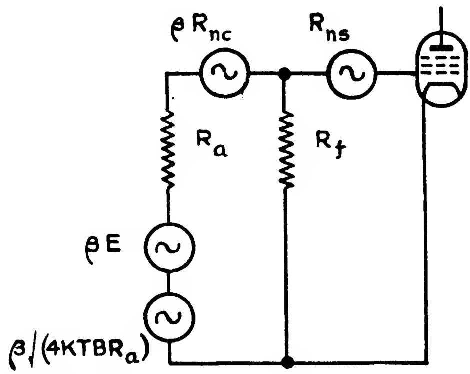


Fig. 3·7. Simplified equivalent circuit of pentode demonstrating effect of positive feed-back on noise factor.

Without additional feed-back

$$N_r = 1 + \frac{R_{nc}}{R_a} + \left(\frac{R_a + R_f}{R_f}\right)^2 \frac{R_{ns}}{R_a},$$

and if feed-back is arranged from anode-plus-screen to grid so that the signal, cathode noise and aerial voltages are increased by the factor β, without affecting the partition noise,

$$N_r = 1 + \frac{R_{nc}}{R_a} + \frac{1}{\beta^2}\left(\frac{R_a + R_f}{R_f}\right)^2 \frac{R_{ns}}{R_a}. \qquad (3·10)$$

It would not need a very high value of β to make the last term insignificant, but the accurate maintenance of even a small amount of positive feed-back is difficult as an engineering proposition.

One alternative possibility for reducing partition noise is the use of negative feed-back from screen grid to control grid.

3·7. Triode radio-frequency amplifiers

From the figures given in § 3·9 it will be seen that most of the noise in pentode valves is due to randomness in the sharing of electrons between the screen and anode, and the importance of this 'partition' noise becomes even greater at high frequencies owing to cathode-lead feed-back, as explained in the previous section. The question which naturally arises is whether it is really necessary to have a screen at all, or in other words, can we use a triode?

The difficulty in using a triode is due, as is well known, to feed-back through the anode-to-grid capacity which normally causes instability if the stage gain is appreciable. Fortunately, there are several ways in which satisfactory amplification with triodes can be obtained at ultra-high frequencies, and the realization of this objective constitutes one of the principal advances made recently in receiver technique.

The first system to be applied was the grounded-grid amplifier, already well known as a transmitting device, in which the signal is applied to the cathode, output taken from the anode, and the grid earthed so that it forms an electrostatic shield between the input and output electrodes. Another arrangement which can be employed, though not so much used for this purpose, is the grounded-anode or cathode-follower circuit which avoids instability by having a voltage gain of less than unity, but provides a power gain in consequence of its high input and low output impedances. The latest device is a return to the earliest of all methods for overcoming the capacity coupling between input and output circuits, namely neutralization, and the neutralized triode is now firmly established as the best arrangement for the first stage of very wide-band amplifiers.

3·7·1. *The grounded-grid triode*

The case of the grounded-grid triode is illustrated by fig. 3·8. From simple valve theory the voltage V_c between cathode and grid causes a current to flow from cathode to anode, given by

$$I = \frac{(\mu+1)V_c}{R_L+\rho}.$$

Hence the input impedance of the valve is given by

$$R_f = \frac{V_c}{I_c} = \frac{R_L+\rho}{\mu+1}, \quad \text{or approximately} \quad \frac{R_L+\rho}{\mu},$$

and the signal voltage at the cathode is given by

$$E_s \frac{R_f}{R_f+R_a}.$$

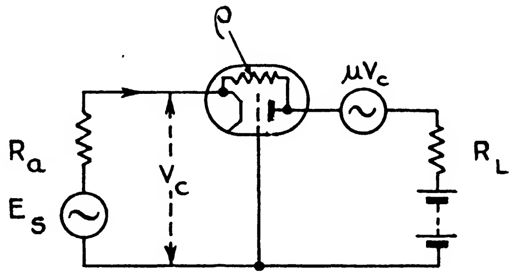


Fig. 3·8. Schematic diagram of grounded-grid triode. ρ is the internal resistance of the valve.

R_n can be replaced by an equivalent anode circuit generator given by $V_n = \mu\sqrt{(4kTBR_n)}$, and with no feed-back this would produce a noise current through the valve given by $i_n = V_n/(R_L+\rho)$; owing to R_a, however, the noise current flowing through the valve will be reduced to some value i_n', in consequence of the voltage $i_n' R_a$ which it produces between grid and cathode. This in turn gives rise to a voltage $(\mu+1)i_n' R_a$ in the anode circuit so that

$$i_n' \simeq i_n - i_n' R_a \left(\frac{\mu+1}{R_L+\rho}\right),$$

and since $\quad \dfrac{\mu+1}{R_L+\rho} = \dfrac{1}{R_f}, \quad i_n' = i_n \dfrac{R_f}{R_a+R_f}.$

The action of feed-back in this case therefore is to reduce signal and noise in the same ratio, the treatment given in the previous section is unchanged, and equations (3·1) or (3·3) may be used for deriving the first-stage noise factor. The effect of feed-back may be represented in the equivalent circuit by a noiseless resistance R_f as shown in fig. 3·9, and the most important practical consequences of such an addition to the circuit are a reduction of gain and an increase in the band-width of the input circuit. This may affect the noise factor indirectly by making it necessary to take account of the second-stage noise, as discussed below. The optimum value of R_a, and therefore the input circuit design, is not directly affected by the presence of R_f; note however that with no feed-back and except at the highest frequencies, satisfaction of equation (3·2) generally leads to the signal source looking into a considerably higher impedance than its own, whereas in the presence of feed-back the situation is reversed, since R_f

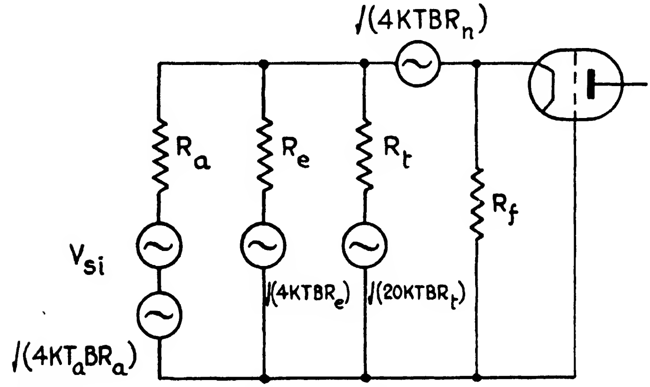


Fig. 3·9. Equivalent circuit for grounded-grid triode, for comparison with figs. 3·1, 3·3 and 3·4. V_{si} is equivalent to E_s.

is normally less than R_a. When the signal is obtained from the output of an efficient mixer, this may modify the standing-wave ratio on the input side of the mixer to some extent; also, in the case, for example, of radar r.f. amplifiers working with a common aerial system, the design of the transmit/receive switch may be influenced.

The noise factor of a triode does not depend directly on the way it is used, since neither R_n nor R_c is thereby affected; on the other hand, the manner of use determines the power gain, and, therefore, the importance of noise from later stages when this is appreciable. This situation may be examined with the assistance of the general theory for networks in cascade outlined in Chapter 2, § 2·10; referring to this, we may choose any arbitrary dividing line between the first stage and the next and assign a noise factor N_1 and gain G_1 to the portion of the receiver up to the dividing line, and noise factor N_2 to the remainder of the receiver, the overall noise factor being given by $N_1+\frac{N_2-1}{G_1}$. In the simplified schematic diagram, fig. 3·10, the dividing line has been placed between the valve and its anode load. If we were to take away everything to the right of the line, so making the valve look into an infinite output load impedance, R_f would be infinite, the input voltage would be E_s (neglecting any reduction due to transit time or circuit losses), and the output voltage, assuming the d.c. anode potential to be maintained somehow, would be

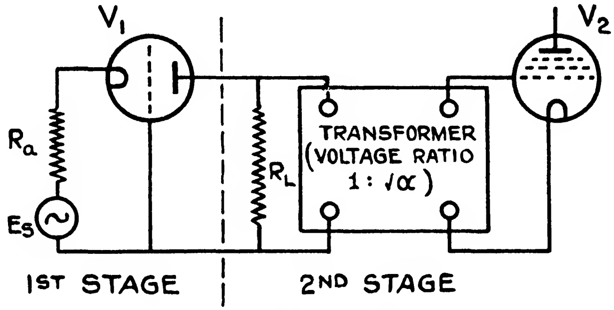


Fig. 3·10. Simplified schematic of grounded-grid triode followed by pentode. (The 'transformer' normally consists of a tapped tuned circuit or a coupled pair.)

μE_s. The output impedance of the valve may be deduced by supposing a voltage E_0 applied to the anode causing a current I_0 to flow through R_a and ρ, and giving rise to an input voltage $I_0 R_a$, which in turn produces an output voltage $-\mu I_0 R_a$. The resultant voltage across $R_a+\rho$ is therefore $E_0-\mu I_0 R_a$, I_0 is given by $(E_0-\mu I_0 R_a)/(R_a+\rho)$ and therefore the output impedance E_0/I_0 is $\rho+(\mu+1)R_a$ or approximately $\rho+\mu R_a$, since μ is large in valves suitable for this purpose. The output power available, at the dotted line in fig. 3·10 and due to the source voltage E_s, is therefore $\frac{\mu^2 E_s^2}{4(\rho+\mu R_a)}$, and the gain G_1 between the terminals of the source and the dotted line is $\frac{\mu^2 R_a}{\rho+\mu R_a}$. With typical valves designed for grounded-grid operation μ is of the order of 100; G_1 can approximate to μ, and will not normally be less than $\frac{1}{2}\mu$, which is its value if the input coupling is adjusted for maximum power transfer when R_L is small compared with ρ, conditions which are likely to exist in very wide-band amplifiers.

With N_2 of the same order as N_1, the $(N_2-1)/G_1$ term in the expression for overall noise factor would be negligible, but the extremely high value of $\rho+\mu R_a$, relative to the anode circuit impedance (consisting of R_L in parallel with $\rho+\mu R_a$) usually demanded by band-width and other practical considerations, means that the part of the circuit to the right of the dotted line is badly mismatched at its input and N_2 is therefore much higher than N_1. N_2 may be evaluated with the assistance of fig. 3·11,

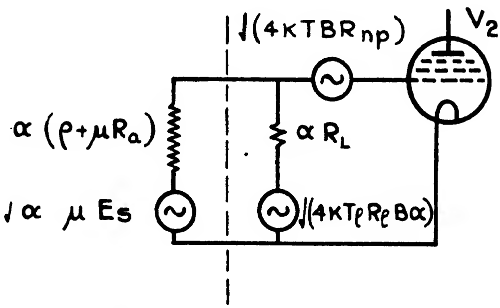


Fig. 3·11. Equivalent circuit showing additional noise introduced by 2nd stage.

which is the equivalent of fig. 3·1 from the point of view of the second valve, V_{si} being replaced by $\mu E_s\sqrt{\alpha}$, R_a by $\alpha(\rho+\mu R_a)$ and R_c by αR_L, which is assumed to have a noise temperature $S_L T$. In this case R_L is a resistance deliberately added, for the purpose of obtaining sufficient band-width, and the required value is given in accordance with the well-known relation between band-width, capacity and parallel resistance, by $R_L \simeq 1/2\pi B(C_a+\alpha C_g)$, where C_a and C_g are the total capacities to ground on the primary and secondary sides of the transformer respectively. It will be assumed that $R_L \ll \rho+\mu R_a$, and since R_L consists mainly of added resistance at room temperature, valve noise may be included in $S_L T$ by the simple relation $S_L \simeq 1+R_{np}/\alpha R_L$.

Proceeding as in Chapter 3, § 3·4, we find

$$N_2 \simeq 1+\frac{S_L}{R_L}(\rho+\mu R_a). \qquad (3·11)$$

The best value for α may be found by substituting

$$1+R_{np}/\alpha R_L \text{ for } S_L \quad \text{and} \quad 1/2\pi B(C_a+\alpha C_g) \text{ for } R_L$$

in (3·11). If α is varied keeping B, C_a, and C_g constant, N_2 is a minimum when

$$\frac{1}{\alpha}=\frac{C_g}{C_a}\sqrt{(1+1/2\pi BC_gR_{np})}. \tag{3·11(a)}$$

Combining 3·11 with the expression for G_1 on p. 37,

$$\frac{N_2-1}{G_1}=\frac{S_L(\rho+\mu R_a)(1+\rho/\mu R_a)}{\mu R_L}, \tag{3·12}$$

and adding this to N_1 as given by equation (3·1), putting $T_a=T$ and $S=T_c/T$,

$$N_r\simeq 1+S\frac{R_a}{R_c}+\frac{R_n}{R_a}\left(1+\frac{R_a}{R_c}\right)^2+\frac{S_L(\rho+\mu R_a)(1+\rho/\mu R_a)}{\mu R_L}. \tag{3·13}$$

By the addition of $(N_2-1)/G_1$ to N_1, the optimum value of R_a is modified to

$$R_{a\,(\text{optimum})}=R_c\sqrt{\frac{R_n+\rho^2S_L/\mu^2R_L}{SR_c+R_n+S_LR_c^2/R_L}}. \tag{3·14}$$

In the case of a grounded-grid triode second stage, the function of R_L is mainly performed by the noiseless feed-back component of the input resistance and $N_2\simeq 1$. The source impedance for the second valve is $\alpha(\rho+\mu R_a)$, and this may be substituted for R_a in the expression for the power gain G_1, and in $(\rho+\mu R_a)$, to obtain respectively the power gain and output impedance of the second stage, the dividing line being between the anode and anode load as before. We find that although the power gain to the anode is nearly μ, the output impedance is approximately $\mu\alpha(\rho+\mu R_a)$, so that the power lost when working into a given small load impedance is $\mu\alpha$ times greater than before.

These results are slightly elaborated, rather than altered appreciably, if account is taken of the various transit time and circuit losses in computing gain, and of second-stage transit time loss in computing N_2. The valves in the two grounded-grid stages have been assumed identical.

3·7·2. *The grounded-anode triode*

Fig. 3·12 is a simplified schematic diagram of a grounded-anode triode circuit, which is to be compared with fig. 3·10. Provided R_t is large compared with the combination of R_a with R_e, in parallel, the output impedance excluding the load resistance R_L is given as $\rho/(\mu+1)$ (or approximately $1/g_m$) by the well-known theory of the cathode follower, and the output voltage is $\simeq E_s$. The power gain is given approximately by $R_a g_m$ and is considerably lower than in the grounded-grid case, but as the output impedance is very low the gain can all be realized in practice if required. It will be noticed that the impedance presented to the valve between grid and cathode depends on R_L, and the equivalent

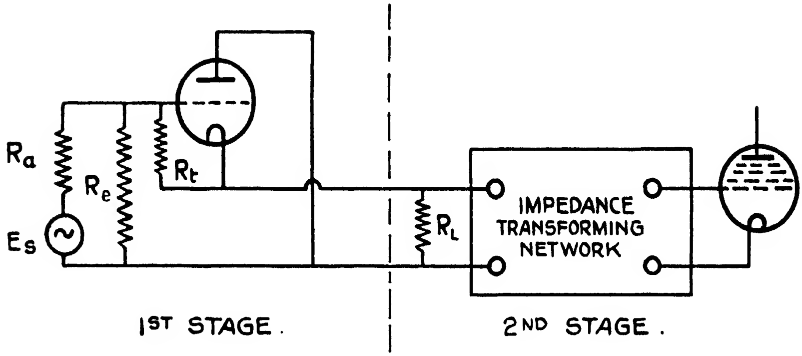


Fig. 3·12. Simplified schematic diagram of grounded-anode triode.

circuit of fig. 3·3 with the analysis based on it is only applicable for small values of R_L. With large values of R_L the signal and thermal noise from R_a is reduced relative to the noise from R_t so that the noise factor becomes rather worse than in the grounded-grid case.

3·7·3. *The neutralized triode*

The advantage obtained from grounded-grid triodes as r.f., and to some extent as i.f. amplifiers, with band-widths of a few megacycles, led to their trial as input stages for band-widths of the order of 20 Mc./sec., but little or no improvement was obtained in comparison with pentodes. The reason for this is the low power gain, a limitation that can be avoided by using a neutralized triode arrangement as shown schematically in fig. 3·13 (which is to be compared with figs. 3·10 and 3·12), the second stage being identical for the purposes of this discussion with that of fig. 3·10.

The available power at the anode of the first valve is $\mu^2 E_s^2/\rho$, and the power gain to the same point $\mu^2 R_a/\rho$. Following the lines of previous reasoning (§ 3·7·1) the noise factor of the second stage is $1+S_L\rho/R_L$, and the contribution of this to the overall noise factor is $S_L\rho^2/\mu^2 R_L R_a$, which is less than that for the grounded-grid case by the factor $1/(1+\mu R_a/\rho)^2$. S_L can be made small by using a grounded-grid second stage, so that wide band-width is obtained without the need for a noisy damping resistance.

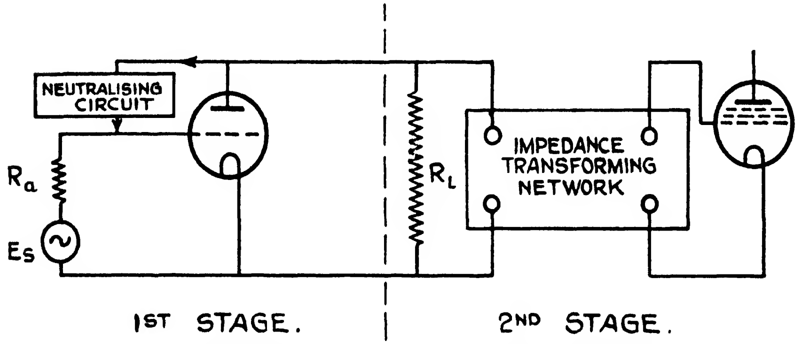


Fig. 3·13. Simplified schematic diagram of neutralized grounded-cathode triode.

It is possible to make a clear distinction between narrow- and wide-band amplifiers according to whether or not the band-width is small enough to allow equation (3·2) to be satisfied. With the majority of triodes suitable for grounded-grid operation there is little to be gained from any increase of power gain at band-widths up to, say, 3 or 4 Mc./sec., but it is worth while at this point to study equation (3·3) with a view to seeing how noise factor might be improved in the 'narrow-band' case. As explained in § 3·5 we can split R_c into the two components R_e and R_t, of which R_t is inversely proportional to frequency squared and R_e more or less inversely proportional to frequency. At the higher frequencies R_t becomes much more important than R_e; the latter can therefore be neglected to a first approximation, R_c becoming the same as R_t. S is now equal to 5, and provided R_n/R_t is appreciably less than 1 (as it usually is in practice over the range of greatest interest) equations (3·3) and (3·9) reduce to

$$N_r = 1 + 2\sqrt{(5R_n/R_t)}. \qquad (3·15)$$

It is well known that for a triode R_n is proportional to $1/g_m$ (§ 3·9) and it has been shown by D. O. North* that R_t is proportional to

* *R.C.A Rev.* Jan. 1942 and April 1940.

$1/G_m\tau^2$, where τ is the transit time. Substituting these results in (3·15) we find

$$N_r = 1 + \theta\tau, \tag{3·16}$$

where θ is some constant, so that for the narrow-band case N_r can be reduced only by reduction of τ and not appreciably in any other way.

By means of neutralization it is possible to avoid the relatively stringent requirements of grounded-grid stages as regards gain and low anode-cathode capacity and the valve can be designed or chosen with a view to minimum transit time alone.

In the case of very wide band-width, such that equations (3·2) and (3·3) are not applicable, noise factor does not depend appreciably on transit time but, to a first approximation, only on R_n and the valve input capacity as explained in the next section.

3·8. Noise factor with very wide band-width

In the case of the grounded-grid triode the band-width of the input circuit is very wide, as pointed out in § 3·7. For large band-widths, however, R_L has to be made rather small and noise from the second stage becomes a serious consideration. Loss also occurs because the input circuit, ignoring R_f (§ 3·7), has a narrow band-width and fails to satisfy the assumption of constant resistive impedance over the pass-band. In the case of a pentode or a neutralized triode, band-width considerations may prevent fulfilment of the optimum input matching condition, and in such cases band-width may be increased without introducing any fresh source of noise, either by using suitable negative feed-back or by reducing R_a. The optimum condition is sufficiently flat to permit a considerable departure without serious loss of performance, as illustrated in fig. 3·14, which shows the variation of N_r with R_a as given by equation (3·5) for the limiting conditions

$$R_f = R_e = \infty \quad \text{and} \quad R_t = R_e = \infty.$$

Curves are also shown for intermediate conditions, based on equation (3·1) putting $T_a = T$ and $S = 1$. Increase of band-width by the alternative method of additional resistance across the valve input produces greater attenuation of the signal relative to valve noise and therefore has a much more serious effect on the noise factor.

3·8·1. *Lower limit of noise factor*

With any given valve and for any given band-width, there is an absolute lower limit to the noise factor obtainable. In the case of an i.f. amplifier, R_c may be increased to a very high value by choosing a sufficiently low intermediate frequency, since the input resistances due to feed-back in the cathode lead and that due to

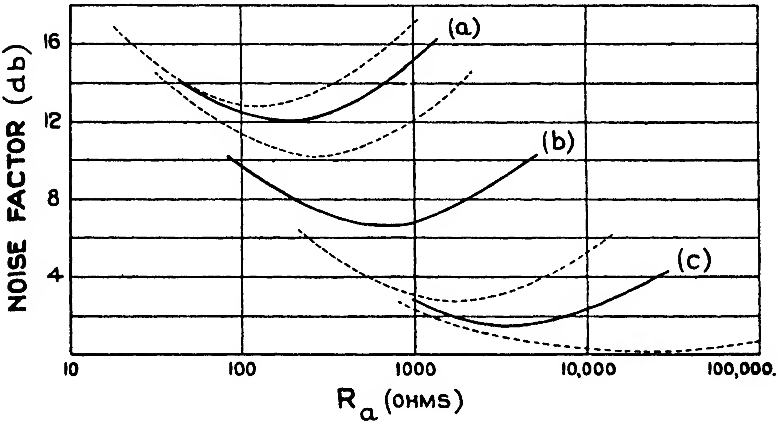


Fig. 3·14. Variation of noise factor with source impedance taking $R_n = 700$ ohms. (*a*) $R_c = 200$ ohms, (*b*) $R_c = 1000$ ohms, (*c*) $R_c = 20{,}000$ ohms. Full lines calculated from equation (3·1) putting $T_a = T$ and $S = 1$, and dotted lines from equation (3·7) for the extreme conditions, i.e. transit time losses only (upper curves) and cathode lead feedback losses only (R_c replaced by R_f).

transit time are both inversely proportional to the square of the frequency. If R_c and R_f are very high, equation (3·5) gives $N_r \simeq 1 + R_n/R_a$, and since the band-width B_t of the input circuit is approximately $1/2\pi CR_a$, where C is the total input circuit capacity, the lowest possible value of first-stage noise factor is given by

$$N_1 = 1 + 2\pi CB_tR_n. \qquad (3·17)$$

This is *not* a function of frequency.

For circuits employing the RL7 pentode, 15 pF. is a reasonable minimum value to take for C, and R_n is about 700 ohms; the minimum value of N_1 is therefore 1·26 (or 1 db.) at 4 Mc./sec. band-width. For an overall band-width of 20 Mc./sec., the input circuit will require to be, say, 30 Mc./sec. wide; then N_1 is 3·0 and practical values for N_2/G_1 and N_3/G_1G_2 are about 1·0 and

0·4 respectively, so that N_r should amount to about 4·4. This is in good agreement with laboratory measurements.

3·8·2. *Upper limit of band-width for satisfying the optimum conditions*

For the case represented in fig. 3·3,

$$\frac{R_a R_c}{R_a+R_c} \simeq \frac{1}{2\pi C B_t},$$

where B_t is the band-width to 3 db. down of the input circuit, assuming this to be a single-tuned circuit. Combining this with equation (3·2) we find that the optimum condition can be satisfied up to an input circuit band-width given by

$$B_t = \frac{1+\sqrt{(1+SR_c/R_n)}}{2\pi R_c C}. \qquad (3\cdot18)$$

Assuming $R_c \gg R_n$, this gives $R_c \simeq S/4\pi^2 C^2 B_t^2 R_n^2$ and (3·3) becomes

$$N_1 \simeq 1+4\pi C B_t R_n. \qquad (3\cdot19)$$

This is considerably worse than the lower limit of noise factor given by (3·17)—1·9 db. instead of 1 db. for the narrow band example—and it is therefore well worth while to increase R_c (e.g. by lowering the i.f.) even when band-width requirements are such that the condition for minimum noise factor can no longer be satisfied.

3·8·3. *Variation of noise factor over the i.f. band-width*

The noise factors so far considered have been the noise factors at the centre of the pass band, since the input impedance of the receiver has been taken as purely resistive. Unless the band-width of the input circuit is wide compared with the overall band-width, the effective noise factor is not necessarily equal to the mid-band noise factor and will tend to be different for small elements of band-width centred at different frequencies within the acceptance band. The worst condition is when the overall band-width is determined entirely by the input circuit. An expression can be simply derived for the '3 db. down' points; referring, for example, to fig. 3·1 a reactance may be introduced in parallel with R_c and of magnitude $(R_c R_a)/(R_c+R_a)$. This reduces the noise and signal voltage across R_a and R_c by 3 db., but leaves the valve noise unchanged so that with $T_a = T_c = T$ equation (3·1) becomes

$$N_r = 1 + \frac{R_a}{R_c} + \frac{2R_n}{R_a}\left(1 + \frac{R_a}{R_c}\right)^2. \qquad (3{\cdot}20)$$

Similarly, equation (3·17) becomes $N_1 = 1 + 4\pi C B_t R_n$, and the noise factor at the '3 db. down' point for the 4 Mc./sec. band-width example given in § 3·8·1 becomes 1·52, or 0·9 db. worse than the mid-band noise factor. With finite values of R_c the effect is less; thus the 1·9 db. figure quoted in § 3·8·2 becomes 2·1 db. at the edge of the band. There will be some additional loss of noise factor due to the reduced gain at the edge of the band, but the integrated effect over the whole band is not likely to have great significance, and this result is supported by experimental evidence for which the author is indebted to N. Houlding. It will be recalled also from § 2·7 that an increase in total noise, due to deterioration of noise factor at the edge of the band, will not necessarily have as much effect on detectability of weak signals as the same increment of noise at band centre.

The band-width of the input circuit must, in general, be wider than the overall band-width required. For some purposes the energy band-width B is of more significance than the more usually employed '3 db. down' band-width, but in amplifiers involving several tuned circuits, the '3 db.' and energy band-widths are generally almost equal. For a single tuned circuit,

$$B/B = \tfrac{1}{2}\pi,$$

and if a single circuit of '3 db.' band-width B_t is superimposed on an ideal rectangular response of the same band-width, it can be shown that the overall energy band-width is $\tfrac{1}{4}\pi$ times that of the rectangular response.

3·9. Equivalent noise resistance of the valve

This is dependent on a number of factors in valve design and is best determined experimentally, but in default of measurement approximate values can be obtained from empirical formulae. The following expression* for valves with oxide-coated cathodes is in agreement with the valve-makers' figures for the RL7 and

* Based on a formula given by Z. Szepesi, *Wireless Eng.* 16 (1939), 67.

EF 50 pentodes and enables these figures to be analysed in terms of partition noise and cathode noise:

$$R_n = \frac{2 \cdot 2}{G_m} \times 10^{-3} + \frac{20}{G_m^2} \frac{I_s I_p}{I_s + I_p} \times 10^{-3}, \qquad (3 \cdot 21)$$

where R_n is in ohms, I_s and I_p are the screen and anode currents respectively in milliamperes and G_m is the mutual conductance in milliamperes/volt. The second term represents the partition noise R_{ns} and disappears in the case of triodes. This gives the figures shown in table 2 for R_n and R_{ns} for representative valves, but it seems probable from experimental evidence that these figures should be increased by up to 50%.

Table 2

Valve	R_n (ohms)	R_{ns} (ohms)
EF 50	1400	1050
RL 7	700	425
954 Acorn	6600	5070
CV 53	430	—
6AK5 (as triode)	430	—
CV 139	220	—

3·10. Summary of practical experience

The desirability of a high ratio of slope to current is evident from equation (3·21). The high-slope pentodes evolved before the war for television receivers were an important advance in this respect, and were employed successfully in certain S.W. broadcasting receivers. The use of these valves ensured that in most practical circumstances the range of broadcast reception was determined by the external noise background and not limited appreciably by the receiver's own contribution to the noise level.

The first notable step forward in the development of r.f. amplifiers for metre wave-lengths was the Acorn pentode, which had the merits of low capacity and a short transit time which more than compensated for the relatively high noise resistance shown in Table 2, and another interesting development was the EF 50 type of high-slope pentode, employing the novel pressed-glass base type of construction which was more robust and easier to make than the Acorn but not quite equal to it in performance at the higher frequencies. Next came the RL 7 (now EF 54), a modified version of the EF 50 using accurate alinement of screen and control grid wires to 'beam' the current on to the anode; by reducing the screen current this arrangement greatly reduced the partition noise, as can be seen from table 2, and a further feature of this valve was the attempt to reduce cathode-lead inductance by bringing out no less than four separate cathode leads. The performance obtained in this way was comparable with that of the Acorn but with the mechanical advantages of the EF 50. The circuit arrangements used with all these valves followed standard practice, the circuit impedances being kept as high as possible by using the valve capacities with a minimum of additional 'strays' as the tuning elements.

With the dawn of what might be called 'noise-factor consciousness' it was realized that there was still a large scope for improvement, and the appearance of the grounded-grid triode was the next step forward. Further improvements were obtained subsequently by reviving the old idea of neutralization, and applying it to metre wave-lengths as described in § 3·7·3.

3·10·1. *Comparison of theory and practice in the case of pentodes*

The difficulties of applying theory to practical cases when dealing with pentodes have been discussed in § 3·6, but it might be expected that equation (3·3) could be made to give some crude sort of approximation to the actual performance, subject to choosing a suitable value of S. Fig. 3·5 shows the theoretical curve for the RL7 obtained from equation (3·3) putting $S=1$ and $R_c=R_eR_l/(R_e+R_l)$, and it is seen to lie roughly half-way between the limiting conditions corresponding to equations (3·8) and (3·9). It has already been suggested that the slightly high position of the practical curve in fig. 3·5, relative to the mean, is of the order to be expected from cold losses, so it is interesting to try the effect of making a reasonable guess at these and putting $R_f=R_t$. The result is shown in fig. 3·15 taking $R_f=R_t=10^7/2f^2$ ohms and assuming the input circuit to have a capacitance of 15 pF. and Q of 100. The experimental values of noise factor from fig. 3·5 are replotted for comparison and found to be in extremely close agreement with the values calculated on the above assumptions. The inference to be drawn from this is that R_f and R_t are in fact of the same order, although the remarkably close agreement between the theoretical and measured curves is probably fortuitous. Fig. 3·15 also shows for the same conditions, the variation of R_a (optimum), T_c, and input-circuit band-width with frequency.

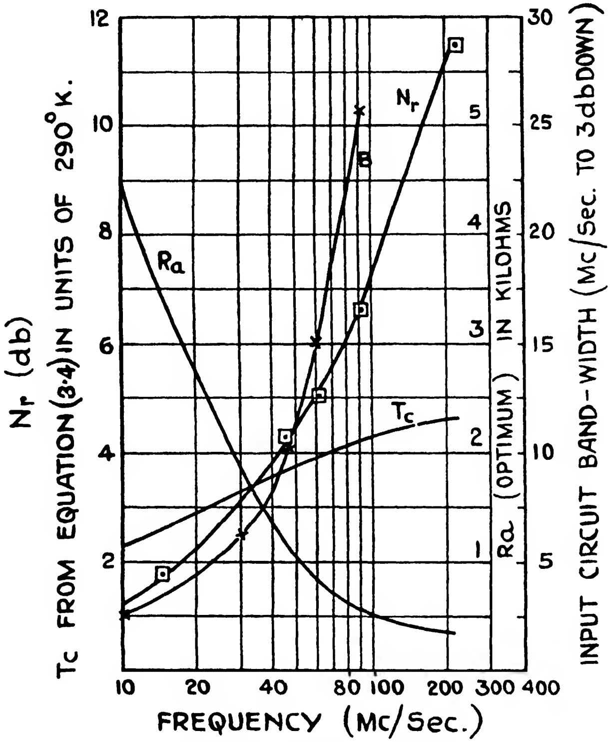


Fig. 3·15. Study of typical RL7 input stage. $R_f = R_t$ and input capacity = 15 $\mu\mu$F. Unloaded Q of input transformer including "cold" valve losses = 100. Points marked ⊡ are experimental values of N_r for typical circuits.

The figure of $10^7/f^2$ for the RL7 is rather worse than the makers' estimate, being based on measurements made on circuits which did not make the best use of the cathode lead arrangement but were typical of the amplifiers used in the experiments. The loss in performance from this cause would not be more than about 1 db. at most.

3·10·2. *Results obtained with grounded-grid triodes*

The first grounded-grid triodes developed for the improvement of receiver noise factor were the S25A and S26A, of disk seal construction and designed by Messrs Standard Telephones and

Cables. The S25A is intended for building into a concentric line assembly working at 600 Mc./sec., and the anode lead consists of a copper tube about $\frac{3}{8}$ in. diameter with a copper-to-glass seal where it enters the bulb. The S26A is similar, but being designed for lower frequencies, employs a conventional type of anode construction with a top cap connexion. Both valves have the following constants:

Internal resistance (ρ)	20,000
Amplification factor (μ)	100
Anode-grid capacity	1·7 pF.
Anode-cathode capacity	0·035 pF.
Cathode-grid capacity	4·0 pF.
Noise resistance (approx. limits)	320–650 ohms

Direct comparison between theory and practice with valves of this type is difficult because they are not suitable for use otherwise than in a grounded-grid circuit, and the transit-time loss resistance is not accurately measurable since it is shunted by the relatively low input impedance due to feed-back. On the other hand, if the theory is firmly established, as seems to be the case, R_n/R_t (equation (3·15)) or θ_T (equation (3·16)) may be deduced from a measurement of noise factor at any one frequency, and the performance at other frequencies predicted by using the relation $R_t \propto 1/\text{frequency}^2$. Fig. 3·16 shows the calculated noise factor of any amplifier in which the input loss is entirely due to transit time, plotted from equation 3·9 as a function of R_t/R_n.

The measured performance of the RL7 (EF54), EF50 and S26A used in typical circuits are all plotted in fig. 3·6 for comparison, and the curve of fig. 3·16 is replotted as the dotted curve in fig. 3·6, fitting it to the measured curve at 214 Mc./sec. (This frequency was chosen because it happened to be a measured point and high enough for the cold losses to be negligible.) Taking R_n from table 2 this makes R_t about 2500 ohms. The slight discrepancy at 600 Mc./sec. between prediction and measurement is well within the probable experimental error, and could alternatively be explained by breakdown of the assumption concerning small transit angles. The discrepancy at the low-frequency end is attributable to two causes, neglect of cold losses and neglect of second-stage noise; thus the cold Q of the input circuit as measured at 60 Mc./sec. was 100, which may be estimated to add 0·6 db., and a further 0·5 db. was the measured contribution from second-stage noise. These figures account for the whole of the difference.

At 200 Mc./sec. a diode mixer having probably about 20 db. noise factor has been used immediately following a single CV 53 r.f. amplifying stage, no appreciable difference in performance for

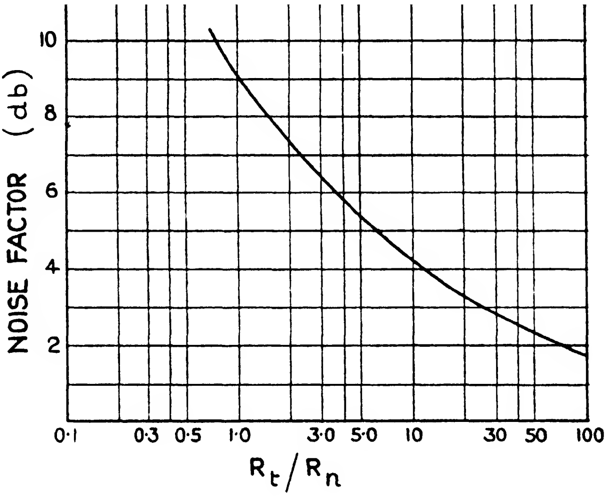


Fig. 3·16. Noise factor of valves with transit time loss onl as calculated from equation (3·9).

band-widths up to 4 Mc./sec. being observed compared with that obtainable with two CV 53 stages, so that second-stage noise can be regarded as negligible, as would be expected on theoretical grounds. An increase of second-stage noise can be expected at lower frequencies on account of the greater input mismatch demanded by equation (3·2) and consequent loss of gain.

The effect of anode-cathode capacity appears to be insignificant under normal working conditions with valves designed for grounded-grid operation, but instability has been encountered with valves not so designed, and is also experienced in amplifiers using the CV 53 if the input and output loading is removed.

3·10·3. *Triodes on pressed-glass bases*

The use of a disk-seal valve is subject to mechanical inconvenience, and grounded-grid triodes on the B9G and miniature bases have been developed for frequencies up to a few hundred Mc./sec. The CV 66 triode on the B9G base has not been investigated in as much detail as the CV53, but at 200 Mc./sec. a noise factor of about 8 db. (or $3\frac{1}{2}$ db. improvement) has been obtained from a single stage preceding on RL7, and about $6\frac{1}{2}$–7 db. from two stages. The lower performance of the CV66 is attributable to the higher capacity, which leads to a much higher value for the term $(N_2 - 1)/G_1$.

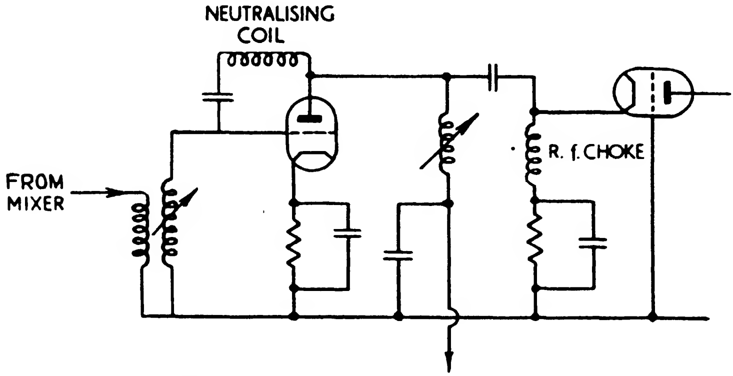


Fig. 3·17. Neutralized triode with grounded grid second stage.

The American 6J4 miniature triode and the roughly equivalent CV139 have been investigated as grounded-grid first i.f. stages in the region of 30–60 Mc./sec. Various measurements have been made both in England and America, and a figure of about 2·8 db. at 45 Mc./sec. for band-widths up to 4 Mc./sec. has been established. Working with appreciably wider band-widths than this, it has been found that second- and third-stage noise is important, and, as explained in § 3·8, little or no advantage has been obtained from the grounded-grid triode in comparison with pentodes.

3·10·4. *The neutralized triode in practice*

A typical circuit arrangement* is shown in fig. 3·17; the anode-grid capacity is tuned by a parallel inductance of low self-capacity, producing a high r.f. impedance from anode to grid. The use of a grounded-grid second stage is advantageous with very high band-widths as explained in § 3·7·3, but with narrow band-widths it can equally well be replaced by a normal pentode.

An i.f. amplifier of this type providing a narrow band-width at

* This is the 'Cascode' circuit described by Wallman, *et al.*, *Proc. I.R.E.* 36 (1948), 700. A noise factor of 0·25 db. at 6 Mc./sec. is claimed, with a 2 Mc./sec. band-width.

45 Mc./sec. and using the CV 139 has been found to give exactly the same performance as a CV 139 grounded-grid amplifier. On the other hand, a triode-connected 6AK5 pentode, which is comparable with the CV 139 in grounded-grid operation, is about 1 db. better as a neutralized triode. The explanation for this, in accordance with § 3·7·3, is the lower transit time of the 6AK5, coupled with insufficient gain for the improvement of overall performance to be realized in the case of grounded-grid operation.

At wide band-widths the difference between neutralized and grounded-grid operation is more marked and less dependent on transit time. As an example, let us work out first the noise factor of a grounded-grid CV 53 amplifier at 45 Mc./sec. having an anode circuit band-width of 30 Mc./sec. (suitable to an overall band-width of, say, 15–20 Mc./sec.) and followed by an RL7 pentode second stage. Reasonable values for C_g and C_a are 15 and 2·5 pF. respectively, and referring to equation (3·11 (*a*)) we find $\alpha = 1/6{\cdot}5$, which in turn makes $R_L = 1100$ ohms. From figs. 3·6 and 3·16 we can deduce that R_t is of the order of 13,000 ohms. The input capacity of the grounded-grid stage is about 7 pF., so that if $Q = 100$ we find that R_e equals 60,000 ohms. This makes $S = 4{\cdot}3$ and $R_c = 10{,}700$ ohms. From the relation $S_L = 1 + R_{np}/\alpha R_L$ we find $S_L \simeq 5$, and substituting these values (including the valve data given in § 3·10·2) in equation (3·14) we find $R_{a\,(\mathrm{optimum})} = 350$ ohms. If this value is substituted in equation (3·13), and if R_n is taken as equal to 430 ohms, then $N_r = 5{\cdot}8$, and a noise factor of this order is obtained in practice.

Referring now to fig. 3·17, the anode load of the second stage must be low to give the required band-width, so its input impedance is approximately ρ/μ. The voltage gain of the first stage is therefore $\frac{\mu}{\rho}\frac{\rho}{\mu}$ or unity, and the effective input capacity C is $C_{gc} + (m+1)\,C_{ga}$, where m is the voltage gain, i.e. $C \simeq C_{gc} + 2C_{ga}$. Stray capacities external to the valve may be neglected as a first approximation and $C_{gc} + 2C_{ga}$ substituted for C in equation (3·17), giving

$$N_1 = 1 + 2\pi B_t R_n (C_{gc} + 2C_{ga}). \tag{3·22}$$

This gives us $R_n(C_{gc} + 2C_{ga})$ as a factor of demerit for the valve, and comparing the 6AK5 with the CV 139 we have $R_n = 430$ and 220 ohms, $C_{gc} + 2C_{ga} = 11$ and 19 pF. and $R_n\,(C_{gc} + 2C_{ga}) = 4750$

and 4200 respectively. Roughly equal performances would therefore be expected from the two valves, and the measured noise factor with 15 Mc./sec. overall band-width at 45 Mc./sec. comes to 4·0 db. in both cases. For an input circuit band-width of 30 Mc./sec. $R_a = 250$ ohms and equation (3·22) gives an average value of 1·85 for N_1.

Following the same procedure as in § 3·7·1 we find, assuming that transit time loss is the principal noisy component of the intervalve coupling impedance, that

$$\frac{N_2 - 1}{G_1} \simeq \frac{5\rho^2}{\mu^2 R_a R_t},$$

and this is obviously small. The next term, $(N_3 - 1)/G_1 G_2$, in the expression for overall noise factor is approximately equal to $S_L \rho^2 / \mu^2 R_a R_L$, assuming that the third stage is a pentode arranged as in fig. 3·10, and taking the same values for S_L, μ, ρ and R_L as in the CV 53 grounded-grid example, this amounts to 0·73 and brings N_r up to 2·58 or 4·1 db. The good agreement with the measured value is somewhat fortuitous as a lot of approximations have been made, including the assumption of a CV 53 second stage instead of the CV 139 used experimentally.

As alternatives to the neutralization scheme shown in fig. 3·17 it is possible to use most of the older arrangements.

3·10·5. *Microwave radio-frequency amplification*

By means of special and rather difficult constructional technique it has been found possible to make triodes and tetrodes with sufficiently small transit time to permit of up to 20 db. amplification at 10 cm. wave-length, but the lowest noise factor so far achieved is several decibels worse than that of the best crystal mixer, and the additional complexity is considerable. Klystrons have also been used, but with no better results because of the difficulty of keeping the beam focused and obtaining sufficiently close coupling between resonators and beam. The most important development for r.f. amplification at microwaves is undoubtedly the travelling wave tube* which has various advantages including easier construction and a band-width of several hundred Mc./sec., although it does not seem particularly promising from the point of view of noise-factor improvement.

* B. J. Kompfner, *Wireless World*, Nov. 1946.

Chapter 4

THE NOISE FACTOR OF MIXERS

4·1. Thermionic mixers

It has been shown in the previous chapter that as the frequency is increased the noise factor of thermionic amplifiers becomes worse on account of electron transit time. The same difficulty applies if we use the valves as mixers instead of amplifiers, with the difference that the noise factor is worse to start with owing to the higher value of equivalent noise resistance due to the fact that the effective mutual conductance (or conversion conductance)

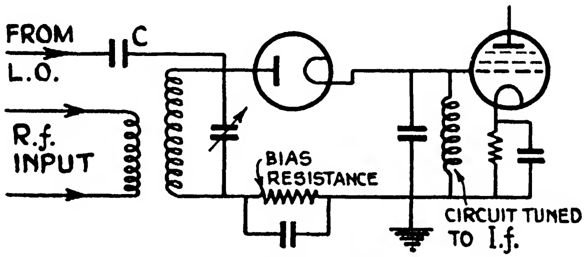


Fig. 4·1. Typical diode mixer.

of a valve as a mixer is in general only about one-third of its mutual conductance as an amplifier, the anode current being about the same in each case; in other words, the signal power output is only one-ninth but the shot noise is the same. With pentagrid and hexode frequency changers the equivalent noise resistance is even worse, because of the larger proportion of the cathode current diverted to auxiliary electrodes, which means, of course, more partition noise.

Rather better performance has been obtained using diode mixers, for which a typical schematic arrangement is shown in fig. 4·1. The condenser *C* provides a large amount of reactive attenuation for the local oscillator, and at the same time prevents loss of signal power into the local oscillator circuit. The i.f. circuit has negligible impedance at signal frequency, and conversely the signal frequency tuning inductance is a short circuit for i.f. The signal must inevitably be attenuated to some extent in the process of frequency conversion, and a certain amount of shot noise is contributed by the diode. At the higher frequencies electron transit-time effects cause additional loss and noise.

At metre wave-lengths the best diode mixers available fall short of the standard set by the grounded-grid triode to the extent of at least 6 db., and it is safe to assume that when low noise factor is important a diode will always be preceded by as much amplification as may be needed to ensure that its contribution to the overall noise factor is negligible. At shorter wave-lengths the performance of thermionic diodes as mixers is surpassed by that of crystal mixers, and at 10 cm. the best diode noise factor so far obtained is about 20 db. compared with 7 db. for the best crystals.

Some attention has been given in the case of centimetre and decimetre wave-lengths to other kinds of mixer, such as those employing velocity modulation or beam deflexion. Hitherto results have not been very promising, the best figures obtained being of the same order as for the diode mixer.*

Thermionic diodes and grid-controlled valves are not usually employed as mixers without r.f. amplification if noise is important.† At microwaves the noise factor is very poor due to the large electron transit angles and in the case of klystron mixers comparable limitations arise from insufficient coupling between electron beams and resonators, and from de-focusing of the beam in the drift space. It is therefore considered that a study of thermionic mixers from the angle of noise factor would not be very profitable, and the remainder of this chapter is written primarily with the crystal mixer in mind. It will be appreciated, however, that the crystal is only one particular kind of diode, and the general treatment given in the next two sections applies equally to other kinds except in so far as it may require to be modified on account of transit time.

* At the moment of going to press, E. W. Herold and C. W. Mueller (*Electronics*, May 1949) announce developments in beam deflexion mixers giving noise factors of the order of 11 db. (average) at 1200 Mc./sec. This is nearly as good as can be achieved with crystal mixers.

† It has been pointed out by M. J. O. Strutt (*Proc. Inst. Radio Engrs, N.Y.*, 34 (1946), 942) that the noise factor of a valve as a mixer can be made an approximation to its noise factor as a triode by suitable use of positive feed-back. The principles involved, and also the practical difficulties, are essentially the same as those which were enumerated (§ 3·6·2) when discussing the use of feed-back to improve pentode noise factor.

4·2 The overall noise-factor of a diode mixer followed by an amplifier

When the first amplifying valve in a receiver is preceded by a diode mixer, the overall noise factor depends on both the mixer and the amplifier, and is assessed most readily with the aid of two new quantities defined as follows and describing respectively the noise and the losses introduced by the mixer.

4·2·1. *Noise temperature ratio (N.T.R.)*

This is defined simply as the ratio of the total noise power available from the mixer at i.f. to the noise power (kTB) available from a resistance at room temperature T.

The N.T.R. is normally greater than unity, but tends to unity as the power injected from the local oscillator tends to zero. Under normal operating conditions the N.T.R. often approaches unity within the usual accuracy of measurement (probably $\pm 0{\cdot}1$), but an N.T.R. of 4 or more is not unusual, and as yet no satisfactory theory has been advanced to account for this enormous variation—at least 30 to 1—in the amount by which the crystal noise exceeds kTB.

4·2·2. *Conversion gain*

Conversion gain is defined as the ratio

$$\frac{\text{power available from the mixer at i.f.}}{\text{power available from the source of signal}}.$$

In accordance with § 2·10, we can divide the receiver into two four-pole networks consisting respectively of the mixer and the i.f. amplifier. By definition of N.T.R. (T_r) and conversion gain (G) the i.f. noise power available from the mixer is $kTBT_r$, of which kTB/G originates in the aerial, so that if N_2 is the i.f. amplifier noise factor, the overall noise factor from equation (2·6) (first two terms) is given by

$$\begin{aligned} N_r &= N_1 + \frac{N_2 - 1}{G} \\ &= T_r/G + (N_2 - 1)/G \\ &= (T_r + N_2 - 1)/G. \end{aligned} \tag{4·1}$$

In practice G is normally less than unity and the term conversion loss ($1/G$) is often preferred.

4·3. Conversion gain

Although the process of frequency changing demands a non-linear characteristic, the output from a mixer varies linearly with the input as long as the applied signal voltage is small enough for the operating point on the mixer valve characteristic to be mainly determined (throughout each cycle) by the local oscillator voltage; also a diode mixer such as that illustrated in fig. 4·1 is capable of

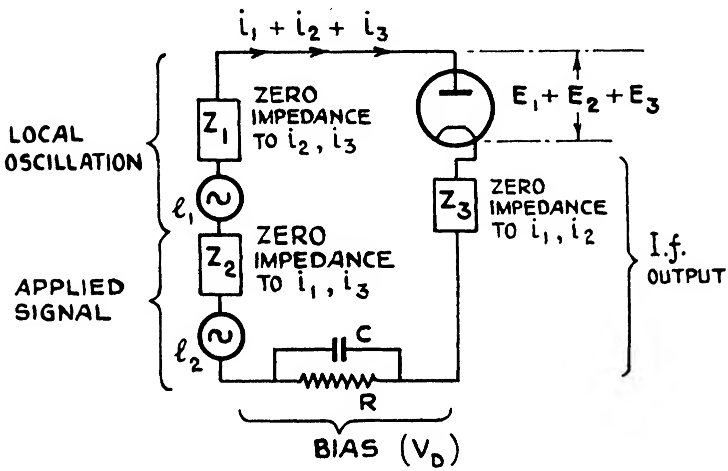


Fig. 4·2. Diagram illustrating currents, voltages and impedances relevant to the performance of the mixer of Fig. 4·1.

transmitting power with equal efficiency in either direction, i.e. from the r.f. to the i.f. side or vice versa. It therefore behaves to a first approximation as a passive linear network, and it can be represented by an equivalent circuit made up of suitable impedances. The precise details of such a circuit depend on what simplifying assumptions can be made, and fig. 4·2 shows the situation as envisaged by E. G. James and J. E. Houldin.*

The voltages applied to the electrodes of the diode are:

$E_1 = (e_1 - i_1 Z_1) \sin \omega_1 t$ at the frequency $(\omega_1/2\pi)$ of the local oscillator,

$E_2 = (e_2 - i_2 Z_2) \sin \omega_2 t$ at the frequency $(\omega_2/2\pi)$ of the signal,

$E_3 = -i_3 Z_3 \sin \omega_3 t$ at i.f. $(\omega_3/2\pi)$,

* 'Diode frequency changers', *Wireless Engr*, 20 (1943), 15.

where i_1, i_2, i_3 are the components of the diode current at oscillator, signal and intermediate frequency and Z_1, Z_2, Z_3, the total external impedance offered to i_1, i_2 and i_3 respectively. It is assumed that transit time is negligible, so that the diode behaves similarly for all frequencies, and that there is no impedance presented to the diode at frequencies other than those considered above. A bias voltage V_D, if required, may either be applied from a battery or produced

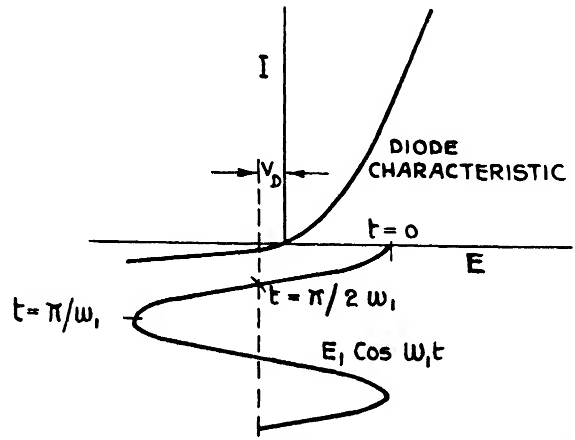


Fig. 4·3. Application of cosine wave to typical diode characteristic.

as shown in the figure by the diode current flowing through the resistance R shunted by the capacity C which acts as a smoothing condenser and simultaneously as a by-pass for i_1, i_2 and i_3.

Fig. 4·3 shows the application of a local oscillation to an arbitrary point on a typical crystal diode characteristic. The instantaneous diode conductance g, which is the slope dI/dE of the current-voltage characteristic at any instant t, varies at local oscillator frequency and can be represented by the following Fourier series:

$$g = g_0 + g_1 \cos \omega_1 t + g_2 \cos 2\omega_1 t + \dots + g_n \cos n\omega_1 t. \qquad (4{\cdot}2)$$

The signal and intermediate frequency voltages add up to $(E_2 \sin \omega_2 t - E_3 \sin \omega_3 t)$ and produce a small change in current given at any time t by

$$\delta i = g(E_2 \sin \omega_2 t - E_3 \sin \omega_3 t) \qquad (4{\cdot}3)$$

Combining equations (4·3) and (4·2) and expanding the product of g and $(E_2 \sin \omega_2 t - E_3 \sin \omega_3 t)$ in terms of sum and difference angles, we find that the amplitudes of the resulting currents at signal and intermediate frequency are given respectively by

$$i_2 = g_0 - \frac{(g_1/2)^2 Z_3}{1 + g_0 Z_3} E_2 \tag{4·4}$$

and

$$i_3 = \frac{g_1}{2} \frac{E_2}{1 + g_0 Z_3}. \tag{4·5}$$

The conversion conductance g_c of any mixer valve may be defined (by analogy with mutual conductance) as

$$\frac{\text{i.f. current in output circuit}}{\text{applied signal frequency voltage}},$$

when the output load impedance tends to zero, and we therefore obtain from equation (4·5) the relation

$$g_c = i_3/E_2 = \tfrac{1}{2} g_1 \tag{4·6}$$

The linear network which is described by equations (4·4) and (4·5) is shown in fig. 4·4 and enables such properties as conversion gain, or the effect of input loading or output admittance and vice

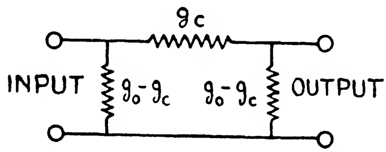


Fig. 4·4. Equivalent circuit of simple diode mixer.

versa, to be predicted from a knowledge of g_0 and g_c. If all losses not connected with the actual process of mixing are represented by an input shunt conductance $\mathbf{G}_1$, the conversion gain M is given* by

$$M = \left(\frac{g_c}{g_0}\right)^2 \frac{1}{[\sqrt{(1 + \mathbf{G}_1/g_0)} + \sqrt{(1 + \mathbf{G}_1/g_0 - g_c^2/g_0^2)}]^2}. \tag{4·7}$$

The determination of g_0 and g_c can be carried out mathematically for idealized characteristics of various kinds, and in the simple case of a linear rectifier having a constant forward conductance g_m and zero conductance in the backward direction,†

$$g_0 = \theta g_m/\pi, \quad g_c = g_m \sin \theta/2\pi, \tag{4·8}$$

* E. G. James, loc. cit.

† Herold, Bush and Ferris, 'The conversion loss of diode mixers having image frequency impedance', *Proc. Inst. Radio Engrs, N.Y.*, Sept. 1945.

where θ is the angle in radians per cycle over which the diode is conducting.

The expansion of equation (4·3) contains terms of frequency $(n\omega_1 \pm \omega_2)/2\pi$, so that a given intermediate frequency $\omega_3/2\pi$ can be obtained not only with a local oscillator frequency $\omega_1/2\pi$ but

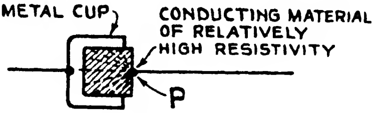


Fig. 4.5. Illustration of bulk resistance. Current paths suggested by dotted lines.

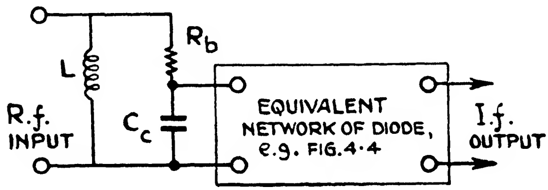


Fig. 4·6. Circuit to illustrate effect of r.f. losses in bulk resistance of crystal.

also with frequencies $\omega_1/2n\pi$, where n is any integer. This is known as harmonic mixing, and proceeding as before the harmonic conversion conductance is given by $\frac{1}{2}g_n$, and for a linear rectifier

$$\tfrac{1}{2}g_n = g_m \sin n\theta/2n\pi. \tag{4·9}$$

In the case of a crystal mixer, the main contribution to $\mathbf{G}_1$ comes from the 'bulk' or 'spreading' resistance of the crystal material. This is nothing to do with the mixing process and arises because in addition to being a rectifier the crystal is also an ohmic resistance having the particular form of construction shown in fig. 4·5. In such a resistance the current spreads out into the bulk of the resistance material from the point P (or alternatively converges on P), and the absorption of power is a maximum in the immediate neighbourhood of P so that there is little advantage in reducing the bulk of the material. The rectification or mixing process takes place in a minute volume at P where conduction is unidirectional, and across this there is a capacity C_c. Fig. 4·6 shows a lumped circuit equivalent to a typical microwave crystal mixer in which R_b represents the bulk resistance, and the external inductance L is arranged to resonate with the capacity C_c. Both

C_c and R_b are measurable and $\mathbf{G}_1$ is given by $\omega^2 C_c^2 R_b$, typical values for C_c and R_b being 0·25 pF. and 25 ohms respectively.

It will be appreciated that, with suitable source and load impedance, power can be transmitted through a network such as that shown in fig. 4·4 with negligible loss provided $(g_0 - g_c)$ is small compared with g_c, and in the case of a linear rectifier this means, as shown by equation (4·8), that sin θ must be nearly equal to θ, i.e. θ must be small. A similar conclusion is reached with other types of diode characteristic such as the $\frac{3}{2}$ power law.* Now when θ is small g_c and g_0 are small, the input impedance is high, and a circuit or bulk resistance loss $\mathbf{G}_1$ which is negligible if θ is large may absorb most of the input power if θ is small. This argument leads to a definite optimum value of θ, tending to zero if $\mathbf{G}_1$ tends to zero.

In practice it has been found with microwave crystal mixers that no improvement is obtained by applying bias; in other words the optimum value of θ cannot be very different from π, and there seems to be nothing much which can be done by the circuit designer to improve performance unless crystals can be developed having much smaller values of $C_c^2 R_b$.

The situation considered above is not an exact representation of typical microwave crystal mixers, because no account has been taken of impedance presented to the rectifying element at image frequency. This impedance is usually appreciable, and its effect has been studied by Herold, Bush and Ferris.† With the addition of Z_4 to represent the total impedance presented to the diode at image frequency $\omega_4/2\pi$, and on the assumption that an instantaneous current i_4 exists at image frequency, fig. 4·2 becomes fig. 4·7. Z_4 and Z_2, although functionally separate, are as a rule the same object physically, namely, the r.f. input circuit of the mixer, and they have also the same electrical value if the band-width of the input circuit is large compared with the separation of signal and image frequencies (twice the i.f.). In place of equation (4·3) we now have

$$\delta i = g(E_2 \sin \omega_2 t - E_3 \sin \omega_3 t - E_4 \sin \omega_4 t) \tag{4·10}$$

Since the signal frequency is given by $\omega_2 = \omega_1 \pm \omega_3$, the image frequency is given by $\omega_4 = \omega_1 \mp \omega_3$ or $\omega_2 \mp 2\omega_3$ or $2\omega_1 - \omega_2$. Note

* E. G. James, loc. cit.

† Loc. cit.

particularly the $2\omega_1$; whereas to obtain all the signal frequency terms in the expansion of equation (4·3) it was sufficient to write

$$g = g_0 + g_1 \cos \omega_1 t,$$

we now have to write

$$g = g_0 + g_1 \cos \omega_1 t + g_2 \cos 2\omega_1 t.$$

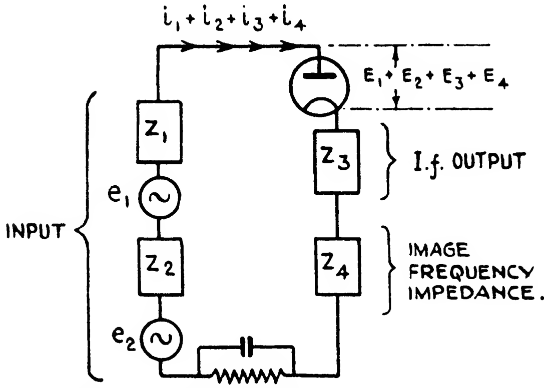


Fig. 4·7. Modification of fig. 4·2 to include image frequency impedance.

Putting $\qquad E_4 = i_4 Z_4$ and $\tfrac{1}{2} g_2 = g_h$,

we have

$$\left.\begin{aligned} i_2 &= E_2 g_0 - E_3 g_c + E_4 g_h, \\ i_3 &= -g_0 E_3 + (E_2 + E_4) g_c, \\ i_4 &= -g_0 E_4 + E_3 g_c - E_2 g_h. \end{aligned}\right\} \qquad (4·11)$$

The equivalent circuit required by equation (4·11) is obviously much more complex than that demanded by equations (4·4) and (4·5). Physical reasoning with fig. 4·4 as a starting point leads to the general arrangement shown in fig. 4·8; this is in exact agreement with equation (4·11) with the values of conductance indicated, and inferences may be drawn from it as follows:

(1) The conversion gain, and the r.f. input and i.f. output impedances will depend upon g_h and on the value of any external impedance at image frequency, i.e. connected to the image frequency terminals of the equivalent network.

(2) The better the match that any such impedance presents to the network at image frequency the more signal power it will absorb, and this loss must be deducted from the power available at i.f.

If the image terminals are short-circuited, fig. 4·8 reduces to fig. 4·4. If there are no circuit losses and g_c is nearly equal to g_0, power can be transferred from the r.f. to the i.f. terminals or vice versa with negligible loss provided the impedances of the source and load are high enough; in other words, the conversion gain tends to unity. On the other hand, in the case of a broadly tuned mixer and a not very high i.f., the image load impedance

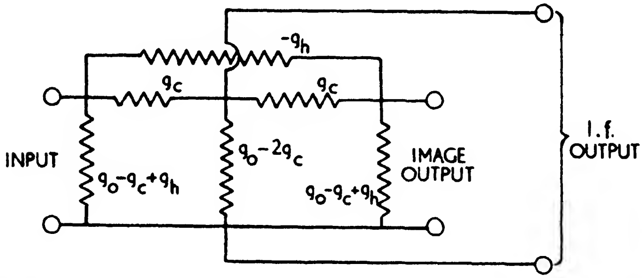


Fig. 4·8. Equivalent circuit of diode mixer, including effect of image frequency impedance.

is identical with the source impedance, and the best result that can be achieved, even with zero loss in the equivalent circuit, is an equal sharing of the available power between the i.f. load and the image-frequency impedance, or in other words a conversion gain of $\frac{1}{2}$, or -3 db. As the values of g_c/g_0 and g_h/g_0 decrease, however, it can be seen that the effect of the load connected to the image-frequency terminal becomes less and less. This is a fairly obvious conclusion, since the image frequency is produced partly as a result of double mixing and is therefore subject to a double conversion loss (the i.f. recombining with the local oscillator to produce image), and partly as a result of harmonic mixing which is in general less efficient than fundamental mixing (e.g. compare equations (4·8) and (4·9)).

Owing to the additional loading and the effect of g_h, one consequence of image impedance may be to increase considerably the optimum value of conducting angle θ. Mathematical analysis of fig. 4·8 is very difficult, but somewhat cumbersome expressions for conversion gain and optimum conditions may be derived from equation (4·11) and will be found in the paper by Herold, Bush and Ferris (loc. cit.).

4·4. Practical crystal mixers

An early type of tunable crystal mixer for 10 cm. wave-length is illustrated diagrammatically in fig. 4·9 (*a*). The length *bc* is about ¼ wave-length, and being earthed at its lower end it presents a high impedance at *b* and can be regarded more or less as an insulating support, disregarding for the moment the local oscillator feed. The length *ab* includes, in series, the R_b and C_c of fig. 4·6

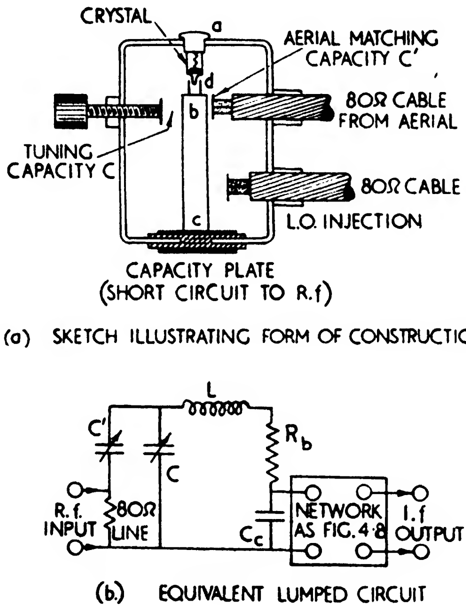


Fig. 4·9. Concentric line mixer. The base capacity is in parallel with the l.f. output and forms the capacity C_s of fig. 3·2 (*c*).

and a certain amount of inductance *L*, the r.f. circuit being completed via the tuning capacity *C*. The feed from the aerial is an 80-ohm line coupled via the capacity *C′* to the lower end of the inductance *L*, and the complete r.f. circuit is therefore as shown in fig. 4·9 (*b*), which is similar in principle to fig. 4·6 with the additional feature that simultaneous adjustment of *C* and *C′* enables the line impedance as seen by the crystal to be varied over a sufficiently wide range to achieve correct matching with the majority of crystals, whilst at the same time maintaining resonance.

To prevent loss of signal power considerable reactive attenuation must be provided in the feed from the local oscillator, as mentioned in § 4·1, and it is physically obvious that if the power from the local oscillator is attenuated, for example, by a factor of 10, one part in ten of the signal power will be lost into the local oscillator. The reactive attenuation for the mixer of fig. 4·9 is provided partly by the capacity between the end of the cable and a point on *bc*, and partly by an adjustable inductive coupling at the other end of the cable to the local oscillator. Some further (mainly resistive) attenuation is provided by *bc* in conjunction with the impedance from *b* to ground, and the slight detuning of *bc* by the local oscillator coupling is compensated by adjustment of *C*. It may be remarked in passing that this arrangement was designed mainly by trial and error, and the explanation of how it worked was not evolved until a long time afterwards!

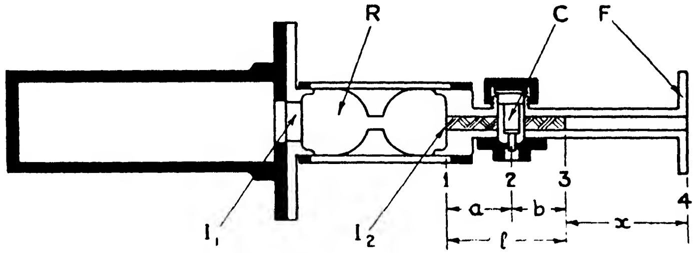


Fig. 4·10. Section of pre-tuned frequency changer, with gas switch and main wave-guide.

A disadvantage of this type of mixer was the critical nature, and interdependence, of all three adjustments—*C*, *C′* and the coupling to the local oscillator—so that maintenance of optimum performance under field conditions was uncertain. The aim of subsequent developments has been to improve reliability, and fig. 4·10 shows a cross-section of a pre-tuned mixer due to H. Dahl* for use in the 10 cm. region. Signal is fed from the wave-guide into the mixer via the resonant cavity *R* which forms part of a gas switch;† the tuning of this is the only adjustment required to the mixer when changing wave-length or inserting a new crystal.

The crystal is placed across a narrow guide which is filled with

* *J. Instn Elect. Engrs*, part III A, no. 1 (1946), 280.
† See Chapter 8, § 8·1.

polystyrene and designed to have a characteristic conductance roughly equal to that of the crystal. The length of the dielectric-filled portion of guide to the right of the crystal is such that it presents at the crystal an inductive reactance, equal to the capacitive reactance of the average crystal. The length of guide to the left of the crystal is therefore matched at the crystal end, and

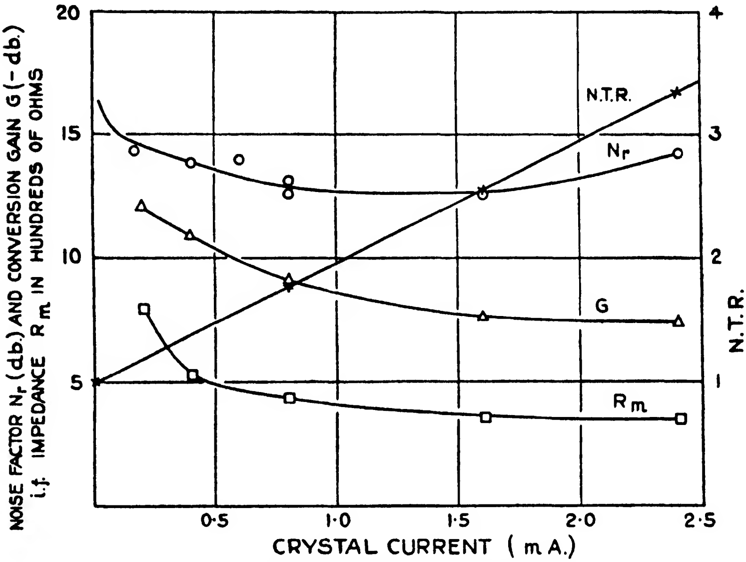


Fig. 4·11. Noise factor, conversion gain, N.T.R. and i.f. impedance measured for typical crystal as a function of crystal current.

since lengths *a* and *b* are both short, the admittance appearing at the junction of the dielectric-filled guide and the resonator has only a small frequency dependence. The coupling of the mixer to the resonator is adjusted by suitable design of aperture, being made sufficiently tight to ensure that the power dissipated in the resonator is small compared with that absorbed by the crystal. The resonator, together with the apertures I_1 and I_2, acts as a complex transformer coupling the mixer to the main wave-guide, and the variations in impedance of production crystals can be compensated to some extent by the tuning of the resonator. Local oscillator power is fed in at point F, and the unfilled portion x of the narrow guide, since it is operating below its cut-off frequency, provides reactive attenuation between the local oscillator and the crystal. The amount of attenuation provided is about 18 db. so that about $1\frac{1}{2}\%$ of the available local oscillator power is supplied to the crystal and, conversely, only $1\frac{1}{2}\%$ of the signal power is fed into the local oscillator and wasted. A frequency band of the order of 4% of the mean frequency is covered by this type of mixer, with a mismatch loss of less than 0·5 db. for the majority of crystals.

The performance of a well-designed microwave crystal mixer depends almost entirely on the crystal, the signal power dissipated in the remainder of the mixer being negligible. The measured performance of a typical crystal is illustrated in fig. 4·11, which shows the variation of N.T.R., conversion gain, noise factor and output impedance with the d.c. crystal current resulting from application of the local oscillator power. During the first few years of microwave radar, crystals were passed into service after testing by means of low-frequency measurements of capacity and rectification efficiency, on the strength of which 'noise factor', or really conversion gain, was predicted, and the acceptance spread was of the order of 4 db. Analysis of the performance of several typical batches of these crystals revealed variations of about 5 or 6 db. in performance, due in more or less equal measure to differences of conversion gain and of N.T.R., the latter becoming relatively more important, and the total spread in performance somewhat greater, with improvement of i.f. noise factor.

One of the difficulties in achieving a consistently high level of performance has been that of testing crystals for N.T.R. under factory conditions, and i.f. amplifier improvements together with the need for closer control of the r.f. parameters of crystals arising out of pre-tuned mixer developments has caused a good deal of effort to be put into the improvement of testing methods. Certain types of crystal are now graded, after manufacture, by insertion in a typical receiver into which an r.f. test signal is fed from a noise source modulated with a square-wave; the percentage variation at modulation frequency of mean output noise level is a measure of the signal-to-noise ratio and is displayed as a reading on a meter which can be calibrated directly in terms of noise factor, since the input signal is automatically maintained at a known and constant level.* In this way the spread in performance from all causes can

* Dresel, Moxon and Schneider, *J. Sci. Inst.* 25 Sept. 1948, p. 295.

be greatly reduced; in tests on a typical batch of 1000 production crystals, just over 20% were found acceptable with a rejection limit only 1½ db. inferior to the best performance, and for many purposes a rejection of four crystals out of five is a small price to pay for the improvement of about 2 db. in average performance.

Equally severe selection methods have been employed in America for some time, with the difference that conversion gain and N.T.R. are measured separately.

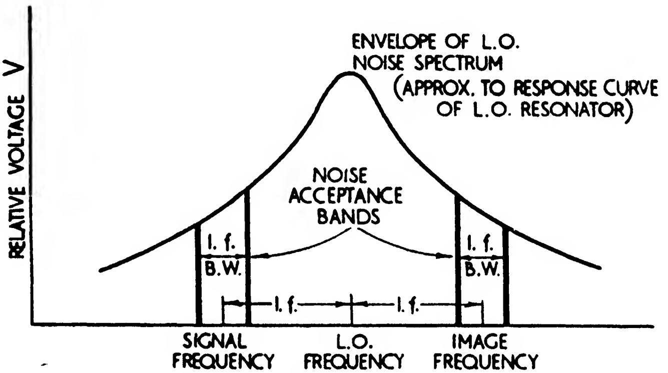


Fig. 4·12. Illustration of L.O. noise and its dependence on the i.f.

4·5. Local oscillator noise

Under certain conditions, common in microwave receivers, the local oscillator of a superheterodyne can make an important contribution to the noise level. The r.f. output of any oscillator is modulated by the shot noise of the valve, and any intermediate-frequency components in the modulation envelope are associated with side-bands at signal and image frequency and converted to i.f. in the mixer. The envelope of the noise spectrum follows in general (though not exactly in all cases) the response curve of the local oscillator resonator, so that, as can be seen from fig. 4·12, the effect is negligible provided the i.f. is high enough in relation to the band-width of the resonator, and as a rule it has only been found serious at microwaves, where r.f. amplification is unsatisfactory and the ratio of i.f. to r.f. is usually rather small. The obvious remedy, i.e. increasing the i.f., is likely to result in increased i.f. amplifier noise (Chapter 3), besides being open to various other objections in many cases (Chapter 6).

Sometimes the resonator of a resistively local oscillator is loaded to overcome a tendency to sudden frequency jumps during adjustment. Such loading by broadening the band-width, increases the response at the signal and image frequencies relative to the mean frequency and therefore results in a larger noise side-band amplitude for a given amount of local oscillator power reaching the mixer, and is to be avoided where possible. In

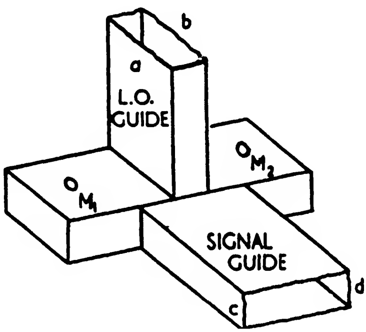


Fig. 4·13. 'Magic-T'.

many cases, however, it is required to use methods of automatic frequency correction (a.f.c.) in which the oscillator frequency is varied electronically; for example (in the case of a klystron) by alteration of reflector potential. This means that the resonator must be capable of supporting oscillations at frequencies of say 10 or 20 Mc./sec. away from its natural resonance without serious loss of efficiency; in other words, it must have a resonance curve 20–40 Mc./sec. wide for a loss of say 3 db. In consequence of the prevailing demand for a.f.c. the problem of oscillator noise has become much more serious than it was in the early days of microwave reception when high-voltage klystrons with high-Q resonators were the fashion.

Of the more obvious remedies, the use of a sharply selective filter to cut out the noise side-bands is ruled out by the a.f.c. requirement and general inconvenience, and rejection of image noise is only a 50 % cure. The trouble can often be minimized if the designer has a choice of local oscillator valves by choosing the most efficient type, because in general if fewer electrons are involved in the generation of a given amount of r.f. power less shot noise is likely to be produced, but this is not an infallible rule.

The best solution is the balanced mixer* in which signal power is fed in the same phase to two mixers, and the local oscillator power in opposite phase, or vice versa. By suitable combination of the i.f. outputs, noise voltages from the local oscillator may be

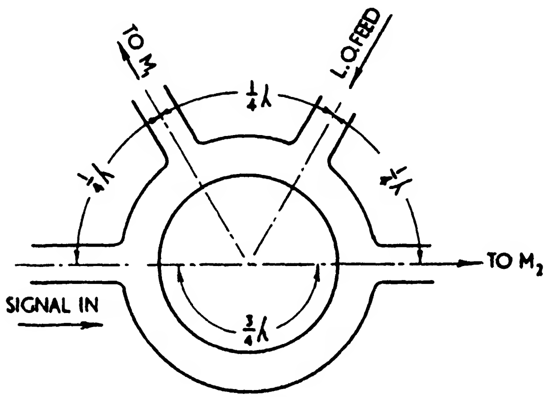


Fig. 4·14. Typical 'rat-race'.

balanced out and signal voltages added. There are several ways in which this procedure may be carried out, but the most usual are the 'magic-*T*' (fig. 4·13) and the 'rat-race' (fig. 4·14). In the magic-*T*, as illustrated, power from the local oscillator reaches the mixers M_1, M_2 in the same phase because the electric fields at the wave-guide faces *a*, *b* are of similar sign, but the signal power is out of phase because the fields at *c* and *d* are of opposite sign. In the particular form of rat-race shown, each mixer receives the signal and local oscillations via alternative paths giving equal phase change, but between the mixers there is a phase difference of 180° for the signal and zero for the local oscillations. The phase relations are preserved in the mixing process, and the addition of signal with cancellation of noise is achieved in both cases by the use of a balanced i.f. input transformer.

Apart from the reduction of noise, there are other advantages

* See, for example, E. G. Schneider, 'Radar', *Proc. Inst. Rad. Engrs*, 34 (1946), 528.

to be gained from balanced mixing; in particular, signals attempting to reach the local oscillator from the two mixers do so in opposite phase and cancel each other so that there is no loss of signal power into the local oscillator and it becomes possible to use much tighter coupling. This means that considerably less local oscillator power is required.

If the crystals in a balanced mixer system are of very unequal performance it is obvious that balancing will not be achieved, and it is interesting to see how critical the balance needs to be. Let the conversion gain of one mixer be G and of the other mG, and let us consider the simple case when the mixers present equal and purely resistive impedances to the source of signal, with correct matching so that each absorbs half the total available signal power P_{si}. If the mixers work into matched i.f. load impedances R and the outputs are connected in series, the total i.f. output power P_{so} is given by [(total output voltage)^2]/$2R$. The voltage outputs from the two mixers are respectively $\sqrt{(\frac{1}{2}P_{si}RG)}$ and $\sqrt{(\frac{1}{2}mP_{si}RG)}$, and since these are added in phase,

$$P_{so}=\frac{[\sqrt{(\frac{1}{2}P_{si}RG)}+\sqrt{(\frac{1}{2}mP_{si}RG)}]^2}{2R}.$$

Similarly, if $\frac{1}{2}P_{ni}$ is the local oscillator noise-input power to each mixer, the noise-output power is given by

$$P_{no}=\frac{[\sqrt{(\frac{1}{2}P_{ni}RG)}-\sqrt{(\frac{1}{2}mP_{ni}RG)}]^2}{2R}.$$

The improvement factor is given by

$$\frac{P_{so}/P_{no}}{P_{si}/P_{ni}}=\left(\frac{1+\sqrt{m}}{1-\sqrt{m}}\right)^2, \tag{4·12}$$

and putting $m=\frac{1}{2}$ or 2 we find that even with a 3 db. difference of conversion gain between the two crystals the improvement is no less than 34 times. With crystals selected to within $1\frac{1}{2}$ db. limits the improvement is not less than 140:1.

Differences of r.f. phase due to the crystal reactances not being fully tuned out by their associated circuits can also cause unbalance and may require either careful crystal selection or the provision of tuning adjustments. The effect of variations in the magnitude of resistive input or output impedance, such as are found in practice, will be small and equivalent to assigning a slightly different value to m in equation (4·12).

It is useful to have some measure of local oscillator noise. It can, of course, be expressed as so much additional noise-input power to the mixer, but it is convenient, particularly when using equation (4·1), to have it described in units of mixer N.T.R. Crystal mixers are usually adjusted so that the local oscillator produces some standard value of crystal current, and it is found that the power required, and therefore the associated noise side-band power, is to a close approximation inversely proportional to the conversion gain G. The available i.f. noise power P_{no} due to the oscillator is proportional to GP_{ni}; P_{ni} is proportional to the receiver band-width and to the percentage noise modulation M of the oscillations, as well as to $1/G$, so that

$$P_{no} \propto G \frac{1}{G} MB,$$

i.e. $P_{no} = AMB$, where A is some constant, and is independent of G. We can therefore express P_{no} as T'/T units of N.T.R. by putting $kT'B = P_{no}$ so that $T'/T = AM/kT$ and is dependent (for a given crystal current) only on the percentage noise modulation of the oscillator and not on any properties of the particular crystal in use.

For the CV 230 (10 cm.) local oscillator, T'/T with 13·5 Mc./sec. i.f. was found to be of the order of 0·1–0·4, being greatest at the edges of the electronic tuning range. With the same i.f. and a 723A (3 cm.) local oscillator, values of T'/T of the order of 1 or 2 have been observed.

Chapter 5

THE MEASUREMENT OF NOISE FACTOR

Methods for measuring noise factor fall into two categories: (*a*) those using a continuous-wave signal generator, and (*b*) those using noise sources.

5·1. Continuous-wave methods

In the continuous wave (c.w.) case a test signal of known available power P_s, is fed into the receiver from an impedance equal to that of the aerial from which the set is intended to work. The detector must be replaced by, or in some way converted into, a device for measuring power ratios.

The required procedure is to adjust the gain of the receiver, with the signal generator off, so that a convenient amount of noise power is indicated. If the signal generator is now switched on and P_s adjusted so that the total power indicated is doubled, the noise factor as defined by equation (2·2) is equal to P_s/kTB. It is not, of course, essential to make the signal power equal to the noise power at the detector; if it is m times the noise power, the following rather more general expression obtains:

$$N_r = P_s/mkTB. \tag{5·1}$$

So much for the principle of the method. Two aspects require further discussion, namely, the determination of m (whether unity or otherwise), and the calibration of the signal generator.

No difficulty arises in determining m if the detector is replaced by a thermocouple or a bolometer bridge, which can be easily calibrated at low frequency, or even d.c., in terms of relative power. This is often rather inconvenient, and it is tempting to try and 'make do' with the detector itself as the power-indicating device. If the detector is a diode, a milliammeter can be connected in series with the load, and in the case of a leaky-grid or anode-bend rectifier a backed-off meter can be used to detect small changes in anode current. Provided the detector is accurately 'square law', i.e. if the current change is proportional to the square of the input voltage, no difficulty arises; a certain meter reading is obtained from noise alone, and $(m+1)$ times this reading is obtained from noise-plus-signal. With other detector laws, however, the detector may behave a little differently according to whether it is fed with c.w. or noise, and this is likely to cause appreciable errors even if the law itself is accurately known. At present the rated accuracy of signal generators, at the low levels required, usually leaves scope for uncertainties to the extent of several decibels, and it was at one time customary to accept a corresponding variation of receiver sensitivity as reasonable, although comparable variations at the transmitting end were usually regarded as serious. When there is no interference present, noise factor is just as important as transmitter power output, a decibel loss in one being exactly equivalent to a decibel loss in the other; this calls for a relatively high degree of measuring accuracy. In achieving a high degree of accuracy with a c.w. signal generator, there are three main problems involved:

(*a*) The signal is necessarily generated at a high level, and elaborate screening and filtering is necessary to prevent stray radiation which can cause appreciable errors, even if not otherwise detectable, by combining in phase or antiphase with the wanted signal.

(*b*) Measurement of signal power must be done at high level because of the insensitivity of measuring devices, and the measured power attenuated by a large amount, say 80–100 db., which must be known to within some fraction of a decibel.

(*c*) The power-measuring device must be calibrated at low frequency and designed so that the frequency error is either negligible or calculable.

Enough has been said to indicate that the problem is rather formidable, at any frequency. It is not insoluble, and even at centimetre wave-lengths a fairly high degree of consistency has been obtained between signal generators having different sorts of power-measuring device and different forms of attenuator.* It can be inferred from this that the degree of absolute accuracy achieved must be fairly high.

5·2. The noise diode

If the c.w. signal generator is replaced by a device which generates a known amount of noise, all the difficulties listed above

* C. W. Oatley, *J. Instn Elect. Engrs*, 93, part IIIA, no. 1 (1946), 201.

disappear. In addition, there is no need to know the energy band-width, and no difficulty arises in allowing for the characteristics of the detector.

The noise source in most general use is the temperature-limited diode in which the anode current has a random or 'noise' component i_n given by $i_n^2 = 2eIB$, where e is the charge on an electron and I is the anode current. If i_n flows through a resistance R, the noise power available from the terminals of the resistance, additional to its thermal noise, is therefore $\frac{1}{2}eIBR$. The impedance of the diode is in parallel with R but is usually of the order of at least 20,000 ohms in normal circumstances. If R corresponds to the aerial impedance, $\frac{1}{2}eIBR$ can be substituted for P_s in equation (5·1), and we find that $N_r = eIR/2mkT$. Taking $T = 290°$ K. and making $m = 1$, we have $e/2mkT = 20{\cdot}0$, and the noise factor is given by

$$N_r = 20IR, \tag{5·2}$$

I being in amperes and R in ohms. Alternatively, expressed in decibels,

$$N_r = 10 \log_{10} 20IR. \tag{5·3}$$

I is adjusted by varying the filament current, and an important practical point to be noted is that the anode voltage must be sufficient to ensure that the current remains temperature-limited even at the maximum value required. Failure to attend to this point usually reveals itself by a decrease of noise with increase of anode current.

Oxide-coated cathodes are unsatisfactory under temperature-limited conditions. Thoriated filaments have been used successfully, but it is safest to employ pure tungsten.

The upper frequency limit for satisfactory operation is set by interelectrode capacity, internal lead lengths, and electron transit time, as explained below.

The circuit arrangement of a typical noise diode assembly using the CV172 is shown in fig. 5·1. An alternative circuit suitable for testing a receiver designed to work from a 400-ohm balanced feeder is shown in fig. 5·2; in this case valve capacity is important because of the higher load impedance, and is tuned out by the inductance L. It has been found most convenient to use a compact diode assembly, separate from its power pack, and to incorporate in the latter the anode current meter and the filament rheostats. It is useful to have two rheostats, one for coarse and one for fine control. The rated maximum anode current of the CV 172 is 30mA. and gives an available power of 24 kTB with 40 ohms resistance, but higher currents can be used for short periods if some reduction of valve life is acceptable. A safe voltage, to ensure saturation up to 50 or 60mA., is 100V.

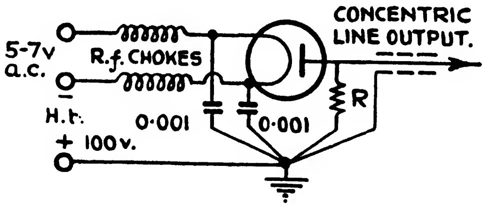


Fig. 5·1. Practical circuit arrangement of noise diode with 40 to 100 Ω concentric line output.

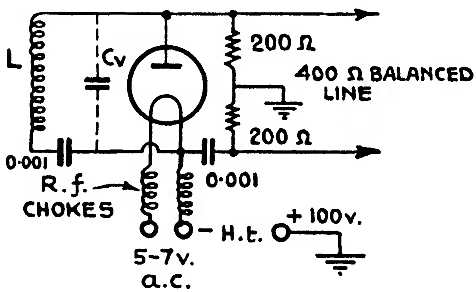


Fig. 5·2. Noise diode for feeding into 400 Ω balanced line.

To determine m, a meter can be used to read detector current, or any other convenient indication of noise power may be used. If the calibration of the indicating device (with its preceding circuits) is unknown, and if sufficient noise power is available, the gain may be turned down so that the receiver noise is negligible, and the noise diode can then be used to find two readings of the meter or other noise indicator which correspond to a 2:1 change of noise power. In practice it is only possible to obtain sufficient diode noise for this procedure if N_r is small, R large, and the required degree of accuracy not very high; for other cases it is necessary to use a slightly more elaborate procedure as follows. With zero current in the noise diode the receiver gain is adjusted to give a convenient meter reading M_1 due to noise. From the definition of noise factor it is clear that this noise level is the same as if the receiver was perfect and the noise power available from the aerial was $N_r kTB$ instead of kTB. The next step is to adjust the noise-diode current to any convenient valve I_1, making available an additional noise power $20IR \times kTB$, and increasing the meter reading to some value M_2 substantially greater than M_1 but taking care to avoid saturation of the receiver. The gain is now turned down appreciably, and the diode currents I_2 (usually made equal to I_1) and I_3 required to give the meter readings M_1 and M_2 respectively are observed. In the first case M_1 and M_2 correspond to a noise-power ratio $\dfrac{N_r}{N_r+20I_1R}$ and in the second case to $\dfrac{N_r+20I_2R}{N_r+20I_3R}$, and since these two ratios are equal we find that

$$N_r = \frac{20\, I_1 I_2 R}{I_3 - I_2 - I_1}. \tag{5·4}$$

Note that if $I_1 + I_2$ is nearly equal to I_3, a small error in readings makes a large error in N_r; this condition should be avoided by making M_2/M_1 as large as possible. It has been assumed that the calibration is not affected by altering the gain; this can be checked by repeating the drill at different gain settings.

5·3. The upper frequency limit for the noise diode

An equivalent circuit for the noise diode is shown in fig. 5·3, in which C_V and L_V represent the interelectrode capacity and lead inductance. Provided the operating frequency is well on the low-frequency side of the resonance of L_V and C_V so that L_V can be ignored, no difficulty arises; it is true that with a large value of R the capacity may exert a shunting effect, but this can be removed by tuning it to resonance with an external inductance such as L in fig. 5·2.

At resonance, if R is small compared with ωL_V, the volts across R are increased by $\omega L_V/R$, i.e. by the 'Q' of the circuit. Since L_V and C_V constitute a series-resonant circuit in parallel with R, the output impedance is merely the series resistance associated with L_V and C_V which may be very small; this means that the receiver is badly mismatched, but the available power is greatly increased. The net result is not amenable to calculation, since L_V and C_V are usually distributed in an obscure manner and not concentrated as shown. The practical importance of the effect is illustrated by the following experience. A CV172 noise diode was being used in the frequency range 200/220 Mc./sec. with a resistance of 40 ohms, and after a time it was suspected that measurements were optimistic to the extent of about 2 db.; by adding an inch of wire in series with the resistance, thus increasing L_V, a further error of 6 or 8 db. in the same direction resulted and the effect described was suspected. Measurement of the total

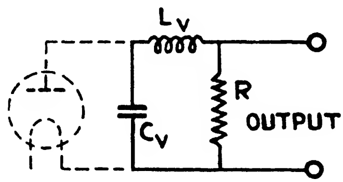


Fig. 5·3. Effect of internal capacitance and inductance of diode.

capacity showed it to be 18 pF.; this was reduced to 5 pF. by removing the valve-holder, the metal base of the valve, and a screen which had been placed over the valve. By adopting a skeleton construction with connexions attached directly to the valve pins and with the assembly attached to and supported by a short length of the 40-ohm cable connecting it to the receiver, the expected difference of 2 db. was obtained. It was later found permissible to replace the valve-holder provided that metal was kept away from the valve and the leads to the resistance and blocking condenser kept as short as possible.

The other source of error which requires discussion is the effect of the transit time of electrons from cathode to anode, which reduces the noise power available at very high frequencies. A test to find out whether this is serious can be made very easily, since the higher the anode voltage the more quickly the electrons are accelerated and the shorter their transit time. It is only necessary, therefore, to change the anode voltage and see what happens; if increased voltage gives increased noise, then the effect is present. With the CV172, no trace of transit-time effect has been observed even at frequencies of over 300 Mc./sec., and the frequency limit for accurate measurements appears to be determined entirely by the resonance effect, which puts it in the region of 250–300 Mc./sec.

5·4. Hot-wire noise source

As an alternative to the noise diode, it is possible to use a metal filament raised to a high temperature.* It is necessary to arrange that this has the resistance value from which the receiver to be tested is designed to work and the noise factor is then given by $(T'-T)/T$, where T' is the temperature of the filament required to double the noise power at the input to the detector.

This method is interesting, but in comparison with the noise diode it has the disadvantages of much less available noise power, and the inconvenience of measuring temperature as compared with measurement of diode current; also, the resistance is usually a function of temperature, and it is therefore necessary to work at a constant temperature to avoid variations of impedance matching.

Very good agreement has been reported between this and the noise diode method at 45 Mc./sec., and there is believed to be no evidence of any serious error at frequencies up to 500 Mc./sec.

5·5. Measurement of aerial noise temperature

It has already been mentioned that the temperature corresponding to the noise power available from an aerial may be widely different from the 290° assumed in the definition of noise factor. It may be very simply measured with a noise diode, as follows. With the aerial connected, a note is made of the output noise reading: the aerial is then replaced by the equivalent resistance and noise diode, and the diode current I required to make the noise level the same as before is observed. By simple reasoning, the amount by which the aerial 'noise temperature' T_a exceeds room temperature T_d is given by $20I_tRT_d$. It is convenient to calibrate the output meter in terms of aerial noise temperature using the relation

$$T_a=(20I_tR+1)T_d, \tag{5·5}$$

and it will be observed that there is no need to have any other knowledge of the detector characteristics. This is the method which was used for obtaining the curves of galactic noise (figs. 2·1 and 2·2).

* E. H. Ullrich and D. C. Rogers, *J. Instn Elect. Engrs*, 93, part III A, no. 1, p. 233.

The method is applicable provided T_a is greater than T_d. It is easily modified for the case when T_a is less than T_d, but this is of relatively little practical importance, since the effect of the value of T_a on the receiver noise output power cannot then be greater than the ratio $(N_r - 1)/N_r$, and at the frequencies to which this applies N_r tends to be relatively large.

If T_a is not large in comparison with T_d, care must be taken to ensure that R is accurately equivalent to the aerial impedance, unless it can be verified that the receiver noise level is not critically dependent on input impedance; otherwise the change of noise level due to any mismatching will be recorded as a 'temperature'.

5·6. Sources of noise at micro-wave-lengths

Klystron noise sources, and specially designed noise diodes, have both been employed for measuring noise factor in the region of 3000–10,000 Mc./sec. So far it has been necessary to calibrate these devices against some other form of signal generator, but the outlook for making absolute measurements with the noise diode up to 3000 Mc./sec. or so is extremely promising.* The diodes used take the form of a tungsten filament mounted centrally in a narrow-bore copper tube about 10 cm. long, and transit time and other correction factors are amenable to calculation. Use has also been made of the hot-wire method at frequencies of the order of 3000 Mc./sec. with some degree of success.

5·6·1. *The double mixing method of noise-factor measurement*

One method of obtaining a microwave noise source is to use an i.f. noise source followed by a mixer which converts the noise to r.f. If this mixer is identical with that in the receiver under test, and if the reciprocity condition holds,† i.e. if the conversion gains from r.f. to i.f. and i.f. to r.f. are the same, then the total gain in db. can be measured at i.f. and divided by two to get the gain in the first mixer. Precautions must be taken to get rid of image frequency noise, e.g. by tuning the local oscillators of the two mixers to

* B. J. Kompfner *et al.* *J. Instn Elect. Engrs*, 93, part III A, no. 1, p. 225.

† Reciprocity is implicit in the representation of mixers by linear networks such as those of figs. 4·4 and 4·8, and is supported by experimental evidence in the case of silicon crystals.

opposite sides of the signal frequency, and some attenuation, of known amount L_a, must be provided between the mixers to avoid undesirable interactions. A low-gain i.f. amplifier is desirable after the diode to make up for some of the losses in the above process.

This method is illustrated in block form by fig. 5·4. Let R_m be the output impedance of the second mixer and let it serve as load for a noise diode D_2 in addition to its normal function. If the noise source on the extreme left is a noise diode D_1, working

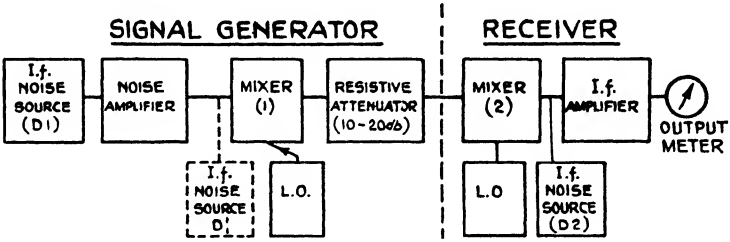


Fig. 5·4. Double mixer method of noise-factor measurement.

into a load R followed by an amplifier of gain G_a, and if I_1, I_2 are the anode currents of D_1, D_2 respectively (only one working at a time) for a given output meter reading, the total conversion gain $G_{m_1}G_{m_2}$ for the two mixers in cascade is obviously

$$\frac{I_2 R_m L_a}{I_1 R G_a}. \tag{5·6}$$

All the quantities on the right are known, or can be determined with very little trouble, and if the two mixers have been carefully matched by a comparative test so that $G_{m_1}=G_{m_2}$, the absolute value of G_{m_1} and G_{m_2} can be evaluated. It is not essential to have matched mixers; we can use any three mixers (or three crystals since, in practice, variations between mixers are almost entirely due to crystals), and the conversion gains G_{m_1}, G_{m_2}, G_{m_3} can be simply evaluated from measurements of $G_{m_1}G_{m_2}$, $G_{m_2}G_{m_3}$ and $G_{m_1}G_{m_3}$.

Knowing the conversion gain G_{m_1} of the first mixer, it follows by further simple reasoning that if I_1 is adjusted so that it doubles the noise power indicated by the output meter

$$N_r = 20 I_1 R G_a G_{m_1} / L_a. \tag{5·7}$$

L_a may be measured simply by insertion and removal, if another attenuator is connected in series with it and provided the source is powerful enough to overcome the total losses.

R_m may be measured in various ways. These include i.f. bridge methods, and deduction from the relationship

$$20 I_2 R_m = (N_{if} + T_r - 1), \tag{5·8}$$

if T_r and N_{if} are known and I_2 doubles the output noise power. If the tuning point of the i.f. amplifier input circuit* is dependent on the value of R_m (as, for example, in the case of fig. 3·2 (c) with $R_s \simeq 1/\omega C_s$), this fact can be used for measurement of R_m by making the inductance variable and providing it with a tuning scale calibrated by substitution of known values of resistance for R_m.

The noise amplifier must have a substantially wider band-width than the i.f. amplifier of the receiver in order to satisfy the requirement that the noise-output power from the receiver shall be proportional to the receiver band-width. A single stage is usually sufficient and although this produces some noise itself, allowance can be made if necessary (i.e. if I_1 is small) by a fixed and easily estimated addition to I_1. In favourable circumstances it is possible to dispense with the amplifier and use the diode D_1' but this must be tuned to prevent the generation of noise at three times the i.f., which (owing to the oscillators being separated by twice the i.f.) produces an r.f. component at the receiver image frequency. Use can also be made of D_1' in checking the amplifier gain.

This method of noise-factor measurement has given good agreement with carefully calibrated signal generators at frequencies up to 10,000 Mc./sec., and retains some, though not all, of the advantages which noise sources offer at low frequencies, namely:

(1) No leakage fields.

(2) Noise factor is determined independently of band-width and of detector law.

(3) There is no r.f. power measurement required.

(4) The need for an attenuator accurate over a range of 80–100 db. is avoided, 10–15 db. being sufficient.

The system has the further convenience that all the components required for it are likely to be at hand in the laboratory, or else are very simply constructed.

* Method due to Dr E. E. Schneider.

5·7. Measurement of conversion gain and noise-temperature ratio

Using the method described above, noise factor is readily analysed into conversion gain and (N.T.R. + i.f. noise factor − 1). A simple determination of i.f. noise factor, replacing R_m with a known resistance and making use of D_2 then enables the N.T.R. to be evaluated.

Other methods for measuring conversion gain include the following:

(1) Direct measurement of the insertion loss of the mixer by use of an i.f. and an r.f. signal generator. This is liable to be inaccurate unless very elaborate precautions are taken in calibrating the generators and avoiding leakage.

(2) An oscillator is modulated at l.f. by a known small percentage, and feeds into the mixer via the normal r.f. input channel with the normal local oscillator switched off or removed. The l.f. output of the mixer is proportional to conversion gain.

(3) By use of the reciprocity principle, the conversion gain may be deduced from the effect on the i.f. output impedance* of known changes in the impedance of the r.f. source.

A common method for the measurement of N.T.R. uses a receiver arrangement such as fig. 3·2 (*c*), or a suitable transmission line, arranged so that the input impedance presented to the first valve, and therefore the receiver noise level, is more or less independent of the value of source impedance over a sufficiently wide range. If, on replacing the crystal by a resistance, the noise power at the receiver output decreases from P_{no} to P'_{no}, we have

$$P_{no}/P'_{no} = (N_{if} + T_r - 1)/N_{if} \tag{5·9}$$

from which we can deduce T_r.

Whatever method is adopted, measurement of noise factor at frequencies higher than 300 Mc./sec. or so is more difficult than at the lower frequencies where diodes such as the CV 172 may be used without frequency changing. On the other hand, measurement of the quantity $(N_{if} + T_r - 1)$ in accordance with equations (5·8) or (5·9) is comparatively simple and tells a large part of the story, particularly as high values of N.T.R. tend to be associated with poor values of conversion gain.

* Method due to R. H. Dicke. For further details see *Crystal Rectifiers*, Radiation Laboratory Series, vol. 15, McGraw Hill Book Co. (London).

Chapter 6

INTERMEDIATE-FREQUENCY AMPLIFIERS

6·1. General

Recent improvement of i.f. amplifier technique has been mainly in relation to the wider band-widths and takes the form of a gradual accumulation of knowledge and experience rather than new ideas.

The majority of i.f. amplifiers for such purposes as television and radar are required to have band-widths of a megacycle or more, and in order to achieve as much gain as possible per stage it is usual to make the circuit impedances as high as possible by using no tuning capacity except the valve interelectrode capacities and other unavoidable 'strays' such as coil self-capacity and wiring capacities. If C is the total capacity of a tuned circuit and B_t is its band-width to 3 db. down, the impedance at resonance is given by the well-known relation $R=1/2\pi CB_t$, and the voltage gain of a simple amplifying stage such as that illustrated in fig. 6·1 is given by*

$$G_v=g_m/2\pi CB_t, \tag{6·1}$$

where g_m is the mutual conductance of the valve. It follows from this expression that there is an upper limit of band-width above which no amplification is possible, and as this limit is approached more and more amplifying stages are necessary to achieve a given overall gain. This situation is in contrast with that obtaining in narrow-band amplifiers, the upper limit for stage gain being then determined usually as a result of feed-back through the grid-anode capacity.

6·2. The product of gain and band-width

With careful design, C is very little greater than the sum C_v of the grid-earth and anode-earth interelectrode capacities, so that equation (6·1) may be rewritten

$$G_vB_t \simeq g_m/2\pi C_v, \tag{6·2}$$

* For this to be strictly true the valve internal impedance must be included in R by using the relation $R = R_2\rho/(R_2 + \rho)$, where R_2 is the parallel impedance of the external circuit, but with pentodes and large band-widths we have $R_2 \ll \rho$.

and the product of gain and band-width* $g_m/2\pi C_v$ is a factor of merit for the type of valve used. It will be noticed that this result is quite independent of intermediate frequency, and although a proper choice of i.f. is important in practical cases, it rests on quite different considerations as discussed later in this chapter.

Although the above definition of factor of merit has been based on a particular case, its use is not restricted to this, since it may be shown for any other specified arrangement (not involving additional capacities) that the product of gain and band-width for a single stage is proportional to $g_m/2\pi C_v$.

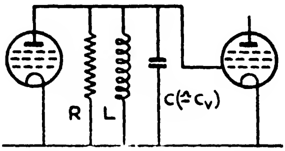


Fig. 6·1. Single amplifier stage (simplified).

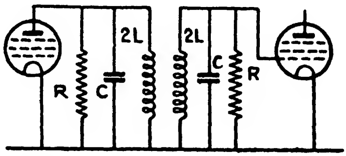


Fig. 6·2. Coupled circuit inter-valve coupling ($C \simeq C_v/2$).

6·2·1. *Coupled circuits*

Consider, for example, a coupled pair of circuits (fig. 6·2); referring to equation (6·4) in the Appendix we observe that the gain at mid-band with critical coupling is exactly half that obtained with a single circuit having the same constants, but since the capacity C is now equal to $\frac{1}{2}C_v$ instead of C_v the gain remains as given by equation (6·1), and is unchanged provided B_t is held constant by doubling R. The stage band-width B_s is considerably greater than B_t, however, so that the product of gain and band-width for the stage is more favourable. The process of increasing circuit complexity to effect a squarer-topped response characteristic could in principle be continued indefinitely, although with very rapidly increasing practical difficulties. It has been shown by W. Hansen† that in the limit the product $G_v B_s$ would be 5·06 times $g_m/2\pi C$, and the frequency response of such a stage would be rectangular, so that any number of stages could be cascaded without any reduction in the product of stage gain G_v and overall

* Throughout this chapter, the term 'band-width' is used meaning the '3 db. down' band-width because this is convenient for calculations, accords with general usage, and usually gives a close approximation to energy band-width.

† *J. Appl. Phys.* 16, no. 9 (Sept. 1945).

band-width B_o. To the engineer endeavouring to meet a given specification with a minimum number of valves it is of course this product $G_v B_o$, and not $G_v B_s$, which is important, $G_v B_o$ being considerably less than $G_v B_s$ in most practical cases.

6·2·2. *Staggered pairs of circuits*

A common method of obtaining a flat-topped response curve is to use simple two-terminal couplings of the fig. 6·1 type, with different stages tuned to different frequencies ('staggered tuning'). It is shown in the Appendix, p. 170, that the amplitude-frequency characteristic for a pair of similar circuits of individual band-width B_t is the same to a first approximation, whether we couple them together with any desired fraction (k) of critical coupling, or separate them (e.g. by a valve) and stagger them by the same fraction (p, equal to k) of their individual band-widths. The same applies to the phase-frequency characteristics. Hansen shows that neither this nor any other type of two-terminal coupling can give a $G_v B_o$ product exceeding twice $g_m/2\pi C_v$. On this basis staggered tuning appears to be much inferior to four-terminal coupling, but in practice there is little to choose between the two methods because in the former case it is practicable to approach much more closely to the ideal. As a typical example, we may compare a two-stage amplifier using critically coupled pairs of circuits with one using a single critically staggered pair. In the first case, the gain per stage is $g_m/2\pi C_v B_t$ as already explained; in the second case, since each circuit is off-time by half its band-width, the gain per stage is $g_m/2\sqrt{2}\,\pi C_v B_t$. The overall band-width for the two stages in the first case is only 80% of that in the second, however, leaving a net advantage of about 15% in favour of the coupled circuits, but even this is more or less wiped out in practice owing to the self-capacity of the additional coils.

6·2·3. *Coupled circuits with unilateral loading*

There are various methods of obtaining an improvement in gain-band-width product relative to that for the simple arrangement considered above without undue complexity. In the case of critical coupling, as may be seen from equations (6·7) and (6·8) in the Appendix, it is possible by transferring all the loading on to one circuit (either primary or secondary), and increasing the coupling between the circuits, to obtain an increase in gain of $\sqrt{2}$ to 1 with no change in frequency response. The response shape under these conditions is much more dependent on accurate adjustment of coupling factor and tuning, and the procedure is only worth while with very wide-band amplifiers, gain being in this case a more urgent requirement and a given percentage tuning error less serious.

6·2·4. *Greater-than-critical coupling*

The use of a coupling somewhat greater than critical is a fairly common practice and often advantageous, especially with a large number of stages. It is found that as the coupling is increased above critical, the product of gain and band-width and the steepness of the sides of the response curve increase steadily. A limit to the possible increase of coupling is set, in the case of most applications, by the appearance of humps in the response curve, but considerable widening of the response occurs before this effect becomes serious. By way of example, fig. 6·3 compares the frequency response, as calculated from equation (6·5)* for five pairs of circuits (*a*) with critical, and (*b*) with 1·2 times critical coupling. The latter shows an increase in overall band-width (to 3 db. down) of 34%, the loss of gain per stage being only $1\frac{1}{2}$%, so that the improvement is nearly as great as with unilateral loading. For a single pair, however, the increase of gain-band-width product would only be about 16%. The improvement is obtained at the cost of somewhat more critical circuit adjustments.

A further improvement, though less than in the case of critical coupling, can be obtained by unilateral loading in addition to overcoupling. From equation (6·12)* this amounts to 1·3 times, and the total improvement compared with simple critical coupling is an increase of about 1·8 times in the value of $G_v B_o$.

6·2·5. *Multiple staggering*

Another method of obtaining an improved gain-band-width product is applicable to two terminal couplings such as fig. 6·1. Staggered circuits may be arranged in similar groups, each embracing two, three, four or more circuits all tuned to different

* Equations 6·3 to 6·21 will be found in the Appendix.

frequencies; the larger the number per group, the better the possible performance but the more critical the adjustments and the greater the dependence on valve uniformity, stability of components, and proper maintenance. The case of staggered pairs has

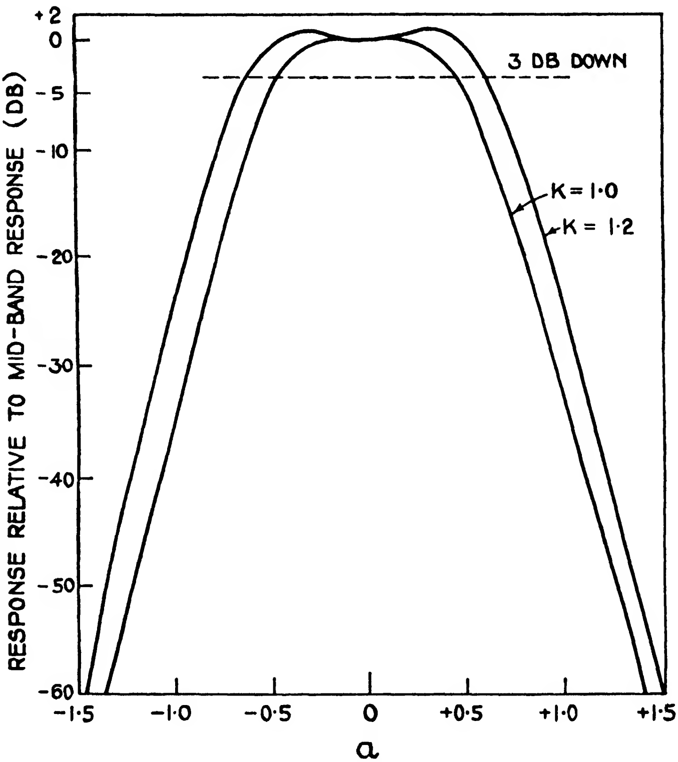


Fig. 6·3. Overall responses of typical amplifiers using 5 pairs of coupled circuits, showing the degree of improvement obtainable by over-coupling; with $K=1{\cdot}2$ the stage (voltage) gain is down by $1\frac{1}{2}\%$ and the overall band-width up by 34%.

been already discussed, and the practical use of the more complex arrangements is limited to very wide band-widths, for exactly the same reasons as in the case of unilateral loading. Any number of such three-circuit groups may be employed, and, in general, the circuits may be arranged in any order and gain control applied to any stage or combination of stages. As with pairs of circuits, increased staggering improves the product of gain and band-width at the price of humps in the response and increased tendency for overshoot, and, in the author's experience, a reasonable compromise is usually achieved by designing for slight humps not exceeding a few per cent. The design of such an amplifier is reasonably simple, as shown in the Appendix, but as the number of circuits per group is increased the calculations rapidly become much more laborious. A very thorough treatment of this subject by H. Wallman* has recently been published, and contains a number of important general conclusions which may be summarized as follows:

(1) If $\mathbf{n}$ stagger-tuned stages are grouped in an optimum arrangement the product of gain and '3 db. down' band-width for the whole group is the same as that obtained with a single stage. Such a group is described as a 'staggered $\mathbf{n}$-uple'. The larger the value of $\mathbf{n}$ the fewer the groups required in cascade and the better therefore the product of stage gain and overall band-width.

(2) With a 'flat-staggered $\mathbf{n}$-uple', i.e. staggering adjusted for maximal flatness, the selectivity function is of the simple form $1/\sqrt{(1+x^{2\mathbf{n}})}$, where x is (for the case of small fractional band-widths) proportional to the departure from the mid-frequency.

(3) The overall band-width of m cascaded $\mathbf{n}$-uples is $\sqrt[2\mathbf{n}]{(2^{1/m}-1)}$ times that of a single $\mathbf{n}$-uple.

(4) When the overall band-width is a large fraction of the mean frequency, staggering requires to be carried out on a basis of geometrical, not arithmetic, symmetry, the circuits of a symmetrical pair having the same Q but different band-widths. The derivation of this result for the simple case of a single pair is given in the Appendix.

The treatment given here, in the Appendix, has been simplified by limiting it to the cases of two and three circuits and the approach is slightly different, the object being to provide the best possible compromise for most practical purposes between adequacy and complexity.

It is well known that the response of a suitable single-tuned circuit may be used to fill in the dip between the humps of a pair of individually narrower but over-coupled circuits, and the latter can, of course, be replaced by an over-staggered pair.

By way of example, fig. 6·4 shows the response as calculated

* See, for example, *Vacuum Tube Amplifiers*, Radiation Laboratory Series, vol. 18, Chap. 4, McGraw Hill Book Co., Inc.

from equation (6·17) for a 15-stage stagger-tuned amplifier employing five groups of three circuits, taking $p=2$ and determining the band-width of the centre-frequency circuits in accordance with equation (6·18). By taking $p=4$ a very slight improvement in $G_v B_o$

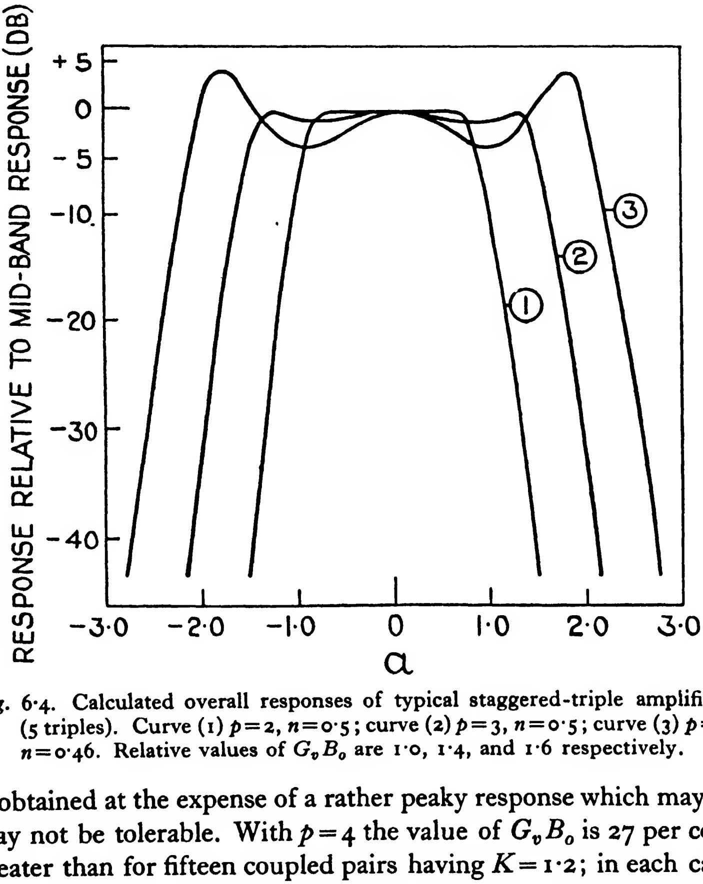


Fig. 6·4. Calculated overall responses of typical staggered-triple amplifiers. (5 triples). Curve (1) $p=2$, $n=0·5$; curve (2) $p=3$, $n=0·5$; curve (3) $p=4$, $n=0·46$. Relative values of $G_v B_o$ are 1·0, 1·4, and 1·6 respectively.

is obtained at the expense of a rather peaky response which may or may not be tolerable. With $p=4$ the value of $G_v B_o$ is 27 per cent greater than for fifteen coupled pairs having $K=1·2$; in each case there are humps of about 3 db. in the response, but in the three-circuit case these could be reduced somewhat by a slight adjustment of centre-circuit band-width.

In the calculation of figs. 6·3 and 6·4 it has been assumed that the ratio of intermediate frequency to band-width is large. When this is not the case, the response is necessarily unsymmetrical, since the curve must cut the X-axis at the points (0, 0) and (0, ∞); it is possible, however, to maintain a level-topped response curve, or one with equal peaks, by suitable design, and in the case of staggered circuits this means parallel loading of the circuits, staggering on the basis of geometrical (not arithmetical) symmetry, and equal power factors (unequal band-width) for the high- and low-frequency circuits, as shown in the Appendix, p. 173.

6·2·6 *Single-tuned circuits tuned to the same frequency*

For band-widths up to about 4 Mc./sec., considerable use has been made of single-circuit couplings all tuned to the same frequency. For a single stage the product of gain and band-width is $g_m/2\pi C_v B_t$ as already stated, but with an increasing number of stages the overall band-width B_o drops very rapidly. Alternatively, if B_o is kept constant the stage gain decreases, and if the process is continued far enough it will be found that the overall gain is actually decreased by adding more stages (Appendix, p. 178).

This arrangement, even for band-widths of a megacycle or so, is costly in valves but has been found attractive in many quarters as it simplifies adjustment and maintenance, very little skill being required to tune any number of circuits all to the same frequency; this is known as 'isochronous' or alternatively as 'synchronous' tuning.

It is interesting to note that in going from the simplest possible arrangement of single circuits (i.e. all tuned to approximately the same frequency) to the most complicated (all circuits tuned to different frequencies) there is a progressive improvement in performance and increase in practical difficulties, and other arrangements such as coupled circuits seem to fall into line more or less according to their complexity.

6·2·7. *Intermediate-frequency amplifiers with negative feed-back*

Intermediate-frequency amplifiers employing negative feed-back have received a good deal of attention since the publication of Wheeler's paper* in 1939. A section of a typical amplifier is shown schematically in fig. 6·5, all the circuits being tuned to the same frequency; such an arrangement is clearly capable of

* *Proc. Instn Radio Engrs, N.Y.*, 27 (1939), 429.

transmitting signals in *either* direction, and it is not surprising to find that it behaves rather like a constant-K filter with the difference that signals travelling from left to right are amplified and any from right to left are attenuated. With a proper termination at the output end of the amplifier, reflexion of the signal can be prevented, and a flat-topped response curve achieved subject to

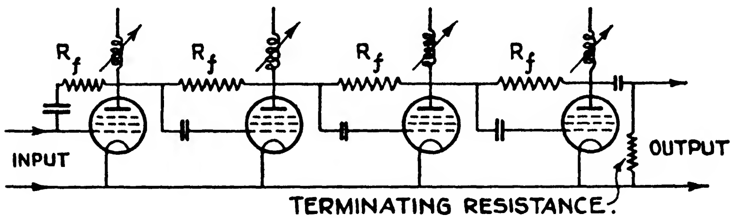


Fig. 6·5. Simplified circuit of negative feed-back amplifier. All circuits are tuned to the same frequency. R_f = feed-back resistances.

suitable choice of component values. Intervalve couplings may be either of the two-terminal type as shown, or four-terminal (e.g. coupled pairs), and although the previously stated limits for gain-band-width product cannot be exceeded, it is found that they can be approached more closely *provided* the stage gain is not too low. This proviso unfortunately means that the advantage is least when most urgently needed, but the usually decisive objection in practice to the employment of negative feed-back is the difficulty of gain control. The amount of feed-back, and therefore the frequency response, is a function of gain, and if this is required to be varied it is necessary to incorporate additional stages of normal type for the purpose. It is usually necessary to provide at least one and sometimes several different kinds of gain adjustment, often with a wide range of control, with the consequence that negative feed-back can at most be employed only to a limited extent and is usually not worth while.

One advantage of negative feed-back i.f. amplifiers, which may be valuable in certain cases, is the relative stability of gain level, the voltage gain per stage being proportional to $\sqrt{g_m}$ instead of g_m.

The product of stage gain and band-width with the arrangement of fig. 6·5 is roughly equal to that of a non-feed-back amplifier using critically coupled pairs. The stage gain is given approximately by $\sqrt{(G_m R_f)}$, and the required terminating impedance is of the order of $\sqrt{(R_f/G_m)}$, although it is not possible to obtain accurate matching throughout the pass-band, and a certain amount of reflexion has to be tolerated.

6·3. Some practical designs of wide-band amplifier

The choice between alternative types of wide-band amplifier circuit depends on the circumstances of individual cases and no general rule can be given.

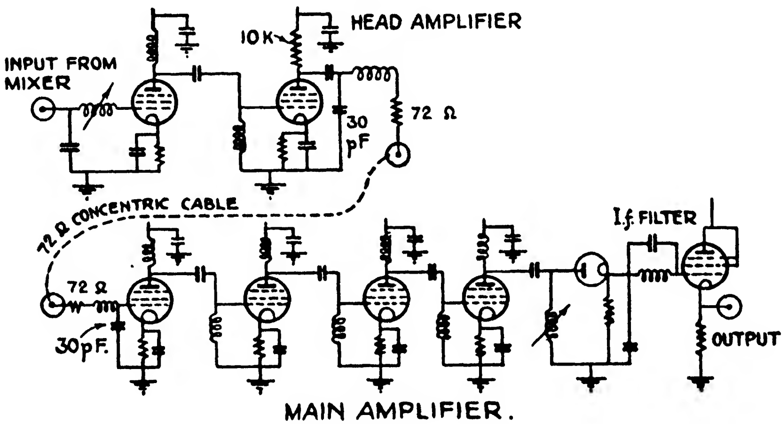


Fig. 6·6. 13·5 Mc./sec. i.f. amplifier with 'top-end' capacity-coupled pairs of circuits.

A typical example of a medium band-width i.f. amplifier system for a radar receiver is illustrated by fig. 6·6. The valves used were type EF 54 for the head amplifier (two stages) and EF 50 for the main amplifier (four stages). The mean frequency was 13·5 Mc./sec., and the coils were wound to have the correct inductance to resonate with the interelectrode capacities, no tuning adjustments being provided except for the first and last circuits. The coupling between grid and anode circuits was 1·2 times critical and obtained by top-end capacity. The head amplifier was coupled to the main i.f. amplifier by a 72-ohm concentric cable terminated at each end by series-tuned circuits having 72 ohms impedance at resonance; from the point of view of gain and band-width this arrangement behaves as a single-tuned circuit intervalve coupling if there is no cable loss, and as two separate circuits of twice the band-width if there is a high cable loss. The additional 30 pF. capacities shown in the diagram were necessary in this instance in order to obtain a sufficiently wide band-width (3–4 Mc./sec.). The measured response was found to be in good agreement with that calculated from the Appendix equations (6·3) and (6·5), in spite of assuming, in order to simplify calculation, that the mean i.f. was large compared with the band-width.

The S26A grounded-grid triode, with its very low capacity (p. 49) and high slope, is an obvious choice of amplifying valve for use when the widest possible band-width is required, and using ten stages with a filter type of interstage coupling a gain of 76 db. with 37 Mc./sec. band-width has been achieved.*

The S26A is mechanically inconvenient to use in this manner, and for band-widths up to 20 Mc./sec. it is usual to employ pentodes such as the 6AK5, EF54 or EF50. EF54 amplifiers have been built using sixteen stages with staggered pairs and having a stage gain of 6 db. with an overall band-width of 15–18 Mc./sec., but this result is easily improved upon by triple staggering or with unilaterally loaded coupled circuits; one amplifier using fifteen EF54 stages with triple staggering had a band-width of 23 Mc./sec. at a geometric mean frequency of 28·2 Mc./sec. and a total gain of 90 db. The outer circuits were tuned to 18·3 and 43·4 Mc./sec. respectively and the ratios of band-width to mean frequency were 0·45 for the outer circuits and 1·0 for the inner circuit.

The electron-multiplier type of valve, though not yet widely employed, is well known as a possible means of achieving a much larger product of gain and band-width. Such valves employ a secondary cathode from which several electrons are emitted, and collected by the anode, for each of the primary electrons. The slope is increased proportionately, without increase of capacity. This type of valve is relatively noisy on account of the randomness of the secondary emission, since it is a matter of chance whether a particular primary electron releases say 2, 3, or 4 secondaries.

* By W. A. Montgomery, as reported by W. B. Lewis, *Proc. Instn Elect. Engrs*, 93, part III A, no. 1, p. 275.

6·4. Distortion in intermediate-frequency amplifiers

The term distortion is applicable to changes other than amplification which take place in the character of signals in their passage through an amplifier. In the case of amplifiers dealing with speech or music, the amplitude-frequency response and linearity of the input-output relationship are the most important characteristics from the point of view of distortionless reproduction, although the phase response is significant in the case of transients. In amplifiers dealing with pulse-type signals, the distortions which have to be considered are as follows:

(1) Alteration in shape of pulse due to imperfections in the amplitude-frequency and phase-frequency characteristics.

(2) Time delay in passage through the amplifier—conveniently considered under this heading but possibly not a 'distortion' in the true sense.

(3) Alteration of pulse shape, and apparent time delay, due to non-linearity in the amplifier.

6·4·1. *Effect of a linear amplifier on pulse shape*

Mathematical investigations into the first of these kinds of distortion usually start by assuming the application of a signal of which the modulation envelope is a Heaviside unit step (fig. 6·7), a pulse being the result of a step-up followed by a step-down or vice versa. It is possible to acquire a useful physical picture of the effect of an i.f. amplifier on pulses without delving any further into the operational calculus, although the latter is the simplest approach for a quantitative treatment.

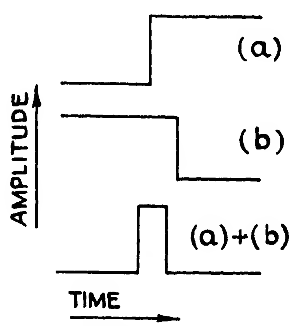


Fig. 6·7. Illustrating construction of pulse from two unit steps.

If the unit step of fig. 6·7 represents the modulation envelope of an r.f. or i.f. signal there must be side-bands corresponding to an infinitely high modulation frequency, because we are concerned with an event which takes place in zero time. Faithful reproduction is only possible therefore with an infinite mean frequency and an infinite band-width B_0. The steepness of rise obviously cannot exceed that of an i.f. cycle, and a little consideration shows that it can only approach this if B_0 is of the order of twice the mean frequency. Usually the band-width is much narrower than this, and the time taken by the output voltage to approach its maximum value occupies several i.f. cycles; similarly, it requires an appreciable time after removal of the excitation for the oscillations to die down. For the simple case of a single-tuned circuit the time constant for the build-up and decay

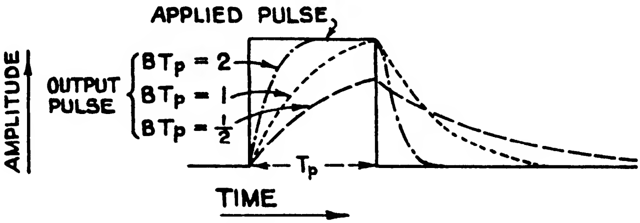


Fig. 6·8. Envelope of oscillations set up in a tuned circuit of band-width B by application of a signal pulse of rectangular modulation envelope for various values of BT_p.

of the oscillations is $2L/R$, and fig. 6·8 shows the response of such a circuit to pulses of various lengths. The time of rise from 10 to 90% of maximum response is given approximately by $0{\cdot}7/B$, and this also holds good for complete amplifiers replacing B by the overall band-width to 3 db. down, B_0.

This does not mean that there is no change of pulse shape as the reponse curve becomes steeper-sided, and fig. 6·9 shows the effect of applying a unit step to an amplifier having sufficiently steep sides. This phenomenon is known as 'overshoot', and the following physical explanation of it—admittedly rather crude—may be found helpful. If we have a sufficiently sharply tuned resonator, such as a tuning fork, and set it suddenly into oscillation, the oscillations persist for a considerable time. Now the response curve of fig. 6·10 has the properties of a sharply tuned circuit from the point of view of the side-bands corresponding to a modulation frequency f_s, and when a unit step is applied we are left after a time with a constant amplitude signal of frequency f_m together with side-bands $f_m \pm f_s$ gradually decaying in amplitude. If at time t these oscillations are all in phase, it can easily be seen that they will be in phase again at times $t+1/f_s$, $t+2/f_s$, etc., and 180° out of phase at times $t+1/2f_s$, $t+3/2f_s$, etc. In other words, the resultant signal amplitude oscillates about the final value with a periodicity of $1/f$. The more steep-sided the response, the more the overshoot, and amplitude response curves of similar shape give the same phase response and overshoot.

The significance or otherwise of overshoot in practical cases is often obscure. The requirements for television are comparatively stringent, and ideally the overshoot should be less than the amount

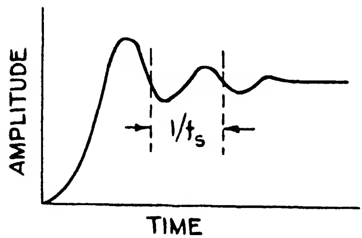


Fig. 6·9. Effect of unit step applied to i.f. amplifier having very steep-sided response curve.

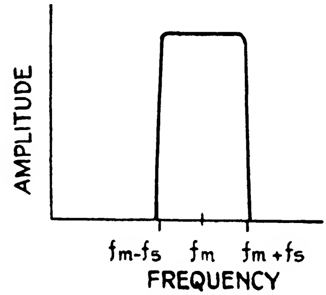


Fig. 6·10. Application of side-band theory to explanation of overshoot.

necessary to give a just perceptible tone gradation. In practice about 5 or 10% is usually acceptable. Overshoot depends of course not only on the receiver response but on the rate of rise or fall of the signal so that it cannot be predicted from receiver characteristics alone.

Typical examples of overshoot with a step-function input (fig. 6·7) are 6·25% for two critically coupled pairs and 17% for three overstaggered triples with 0·1 db. dips per triple.* With isochronous circuits, on tune to the signal, there is no overshoot.

6·4·2. *Time delay*

If there was no time delay in an i.f. amplifier the phase shift would be zero at all frequencies. On the other hand, given a finite time delay τ, a shift $d\omega/2\pi$ in frequency will produce a phase change $d\phi=\tau d\omega$, i.e. $\tau=d\phi/d\omega$. As shown in the Appendix, p. 172, this gives a delay time $1/\pi B_t$ for a single circuit and $2/\pi B_t(1+k^2)$ or $2/\pi B_t(1+p^2)$ for each coupled or staggered pair.

* H. Wallman, loc. cit.

These formulae give, approximately, the time delay as measured to the centre of the leading edge of the output pulse. $d\phi/d\omega$ is, of course, a function of frequency, and by way of example the phase response and time delay corresponding to the amplitude response curves of fig. 6·3 have been plotted in fig. 6·11.

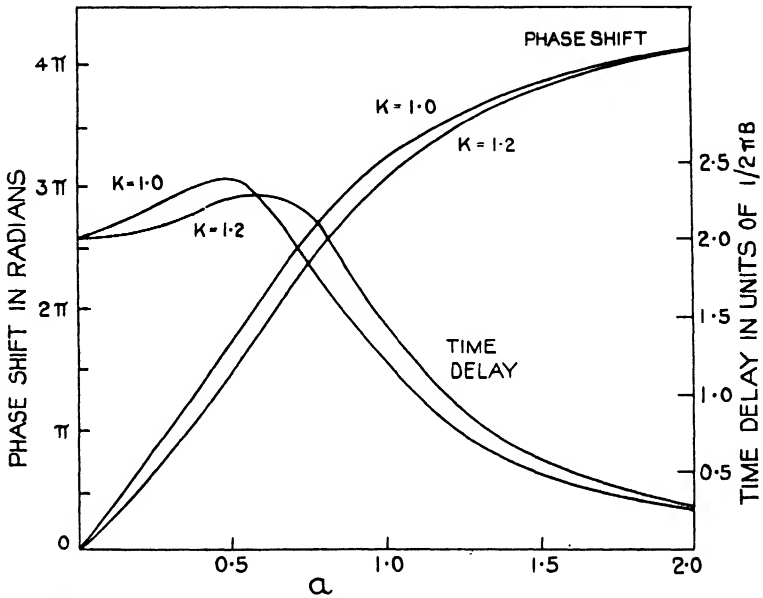


Fig. 6·11. Phase shift and time delay of typical i.f. amplifiers. (These correspond to the amplitude response curves of fig. 6·3.)

Delay does not in itself affect the pulse shape, but different delays at different side-band frequencies obviously must do so. This effect, like the related effect of overshoot, is more pronounced the steeper the sides of the response curve.

In estimating either time delay or overshoot it is of course the overall response of r.f., i.f., and video circuits which is significant. The time delay is (approximately) the time interval between the centre of the leading edges of the input and output pulses.

6·4·3. *Number of cycles per pulse*

When the pulse length is not much greater than the reciprocal of the i.f., so that each pulse contains only a very small number of i.f. cycles, additional distortion of pulse shape arises for two reasons. As already mentioned, it is an obvious physical impossibility for the envelope of the leading edge of the pulse to rise more steeply than the i.f. wave-form. In addition, there is difficulty with the integrating capacity at the output of the detector because if this is small enough to be charged quickly at the start of the pulse and loose its charge quickly at the end, it will be too small to hold more than part of its charge from one half cycle to the next. This latter difficulty can be minimized by the employment of full-wave rectification at the second detector.

6·4·4. *Non-linearity in the amplifier*

In the case of telephony reception, non-linearity in an i.f. or r.f. amplifier has the effect of increasing or decreasing the apparent modulation percentage, and also distorts the envelope of the modulation; these effects are well known and have been adequately covered in previous literature. For pulse reception we require a slightly different angle of approach, and fig. 6·12 shows the effect of a typical non-linear characteristic on trapezoidal pulses of different magnitudes. The main consequences are:

(*a*) A difference in the apparent arrival time for small and large signals, taking any point on the leading edge as reference. (Points *a* and *b* are unusable, *a* being masked by noise and *b* rounded and indefinite in practical cases.)

(*b*) An exaggeration of amplitude differences; the nominal bearing discrimination of a radar system is thereby increased, but fading effects are more pronounced.

(*c*) The range of input which can be handled at a given gain setting without overloading is reduced.

These considerations hold for any non-rectangular pulse shape. In the case of rectangular pulses, the shape remains a rectangle whatever the input-output curve, but (*b*) and (*c*) still apply.

The situation envisaged in fig. 6·12 is one which frequently arises in radar equipments, especially when the gain of high-slope amplifying valves is controlled by alteration of control-grid bias or screen-grid voltage. The importance of this form of distortion depends on the accuracy of range determination required. For very accurate ranging the effective time delay in the amplifier must be constant so that a simple correction can be made for it in the setting up of the range scales, and the extra 'delay' of weak signals as illustrated in fig. 6·12 is not acceptable. If the transmitted pulse is sufficiently steep-sided, the difficulty can be avoided with some sacrifice of signal-to-noise ratio by having

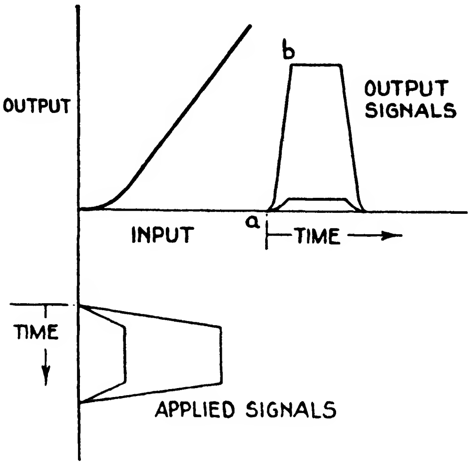


Fig. 6·12 (*a*) Illustration of effect of non-linearity of valve characteristics at high levels in i.f. or video amplifiers.

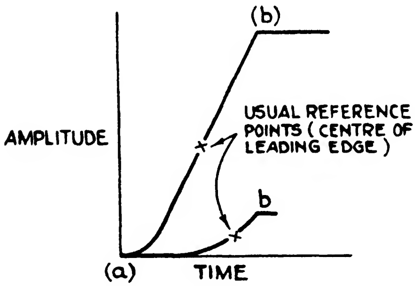


Fig. 6·12 (*b*) Enlarged picture of leading edges of output signals.

Note how any reference point other than (*a*) or (*b*) (which are unsuitable because (*a*) is down in the noise level and (*b*) is normally rounded and indefinite) depends for its timing on signal amplitude.

sufficient receiver band-width; in this way an approximation is obtained to the case of a rectangular pulse. Alternatively, a method of gain control must be used which does not introduce appreciable non-linearity, such as the application of bias to suppressor grids. In using this, care is necessary in two respects; the application of negative bias to the suppressor reduces gain by diversion of electrons from the anode to the screen, so that in order to keep within the rated screen dissipation it may be necessary to operate with a somewhat higher grid-bias voltage or to use a voltage-dropping resistance in series with the screen, and further it is found that the saturation output level is reduced unless the last two or three stages are uncontrolled.

6·5. Overloading and paralysis

Fig. 6·13 shows a circuit diagram for a hypothetical radar receiver of early design. Under ideal conditions such a receiver would be perfectly satisfactory, but it would probably be put out of action by quite weak c.w. or m.c.w. interference, and a block of 'clutter' due to returns from reflecting objects of no interest would be followed by paralysis of the form sketched in fig. 8·4 (*b*). These are consequences of a number of features now established as 'bad practice' in receivers required to deal with pulse-type signals.

Refer to fig. 6·13, and consider what happens in the presence of a continuous signal, or a signal present for an appreciable fraction of the total time, and strong enough to cause one or more i.f. stages to draw increased anode current. The h.t. supply voltage and therefore the screen voltages on *all* the valves, falls and causes a reduction in gain. In the event of the interference ceasing, C_1 has to charge up again before the gain is normal, and as the time constant of R_1C_1 is 4 msec., full recovery may not occur for several cycles of the repetition frequency, by which time the interference has perhaps started again so that the opportunity of acquiring useful information has been lost. The same sort of thing happens in the case of R_2C_2 and R_3C_3, although the time constants in this case happen to be shorter and the result may not be serious except in the case of interference locked to the time-base, such as the radar ground wave or clutter, which gives rise to paralysis of the type shown in fig. 8·4 (*b*). The aim should be to reduce the coupling and the decoupling time constants, R_2C_2 and R_3C_3 respectively, to something shorter than the pulse length. The poor regulation of the power supply, due in this case to R_1, is best overcome by the use of a stabilized power pack, and so far as possible this should be designed to cope with any likely modulation frequencies of the interfering signal.

Other faults remain. If the interference is strong enough,

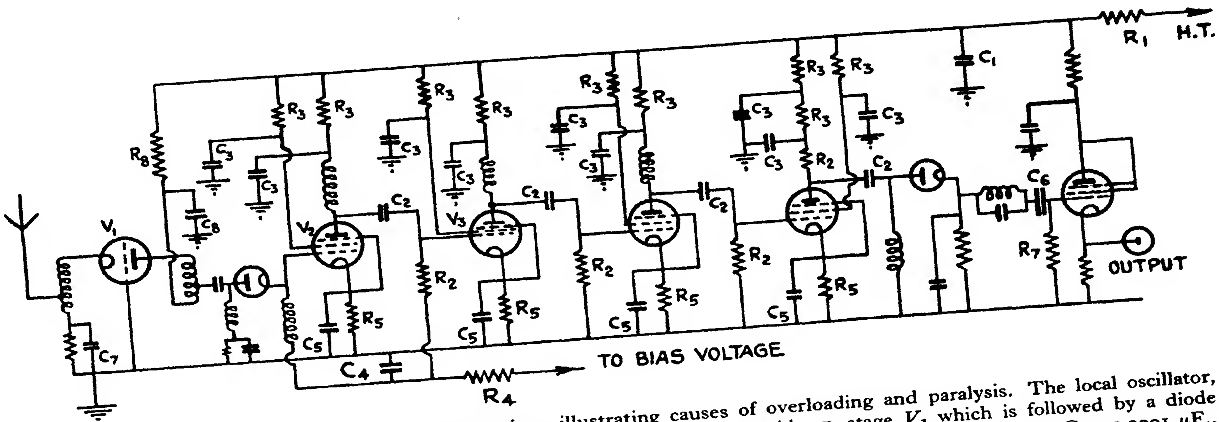


Fig. 6·13. Diagram of early type of radar receiver, illustrating causes of overloading and paralysis. The local oscillator, not shown, is loosely coupled to the anode circuit of the grounded-grid R.F. stage V_1 which is followed by a diode frequency changer. $C_1=4\ \mu F.$, $C_2=0{\cdot}001$, $C_3=0{\cdot}002$, $C_4=25\ \mu F.$, $C_5=0{\cdot}025\ \mu F.$, $C_6=0{\cdot}01$, C_7, $C_8=0{\cdot}0001\ \mu F.$, $R=1000\ \Omega$, $R_2=5000\ \Omega$, $R_3=3000\ \Omega$, $R_4=1000\ \Omega$, $R_5=200\ \Omega$, $R_7=\frac{1}{2}\ M\Omega$, $R_3=3K\Omega$.

V_3 may be driven into grid current. This charges C_4, increases the grid bias, and reduces the gain. The remedy for this is to use either a bias supply of low impedance or to control the gain by bias on the suppressor grid. The time constants R_5C_5 of the cathode-bias circuits can cause the same trouble as R_2C_2, R_3C_3, and the remedy is to reduce C_5, but with low intermediate frequencies it may sometimes be necessary to tolerate a small amount of paralysis in order to obtain sufficient by-passing of the intermediate frequency.

In the cathode-follower circuit, the condenser C_6 gradually charges up to the mean level of any interference so that on the cessation of this interference the cathode-follower valve may be cut off, until C_6 has had time to discharge via the diode and R_7. The effect of this, with the time constants as shown, will be a general reduction of gain, or even complete cut-off, for several traces. A d.c. coupling from second detector to cathode-follower is therefore preferable to resistance-capacity coupling.

Another defect of the circuit of fig. 6·13 is susceptibility to interference at i.f. which arises because the impedance of C_7 and C_8 is by no means negligible at i.f. One of the several possible remedies is to place the bias resistance and condenser at the other end of the input coil. I.f. reaching V_1 can also cause trouble by beating with stray local oscillator voltage which converts it to r.f.

6·6. Feed-back in intermediate-frequency amplifiers

Another important aspect of the design of i.f. amplifiers is the avoidance of feed-back. If feed-back of any sort, positive or negative, is present, it is invariably a function of frequency; and if a gain-controlled stage is included in the feed-back loop this makes the frequency response a function of gain setting as already discussed for the case of negative feed-back. Positive feed-back is in general difficult to control, and leads to peaky i.f. responses and in extreme cases to self-oscillation.

A typical i.f. amplifier may have a band-width of 4 Mc./sec. and a noise factor of two times with a source impedance as presented to the first valve of say 1000 ohms. The effective r.m.s. noise-input voltage at the grid, equal to $\sqrt{(4kTBR_aN_r)}$, is therefore $11{\cdot}2\mu$V., and the required output-noise level may be about 1 V. r.m.s., so that the gain is about 10^5 times. If the allowable feed-back is 10%, this means that not more than one-millionth part of the output circuit impedance may be common to the input circuit, and care in wiring is necessary in order to minimize interaction of i.f. currents circulating in the chassis; in one very badly arranged amplifier the input and output stages were adjacent and self-oscillation took place, but was completely cured by slotting the chassis between the first and last stages, so preventing inter-linkage of their associated 'earth return' paths.

Decoupling and screening can usually be quite simple and high-gain amplifiers have been constructed with *no interstage screening at all.* This result is not so surprising as it may seem, since the chassis with its cover can be regarded as a wave-guide operating a long way below its cut-off frequency.

Another main route for unwanted feed-back is the valve heater supply, and the use of decoupling chokes separating the heaters into at least two groups is usually necessary. Efficient filtering of i.f. at the detector output is also required, in many instances, to prevent feed-back of i.f. from video or l.f. circuitry into the amplifier input.

Subject, therefore, to quite simple precautions, gains of the order of 100 db. can be achieved satisfactorily, but failure to appreciate the essential requirements has sometimes led to the employment of complicated designs, including double frequency changing in order to avoid having a lot of gain all at one frequency.

6·7. Choosing the intermediate frequency

Some of the considerations which affect the choice of i.f. are quite general, and, as they are well known, need not be discussed here in detail; thus the higher the i.f. the easier it is to deal with images and most of the other unwanted responses, but if the i.f. is too high there may be practical difficulties in obtaining sharp enough circuits, or in maintaining the adjustments to a sufficient degree of accuracy under other than laboratory conditions. Difficulties arising when the i.f. is not sufficiently large compared with the band-width have already been discussed (§ 6·4·3). Further considerations likely to affect the choice of i.f. include the requirement of avoiding interference pick-up from, say, a nearby transmitter operating in the region of the i.f., or of avoiding some image frequency or other spurious response.

In receivers for microwaves, additional considerations include the increase of i.f. noise factor at the higher intermediate frequencies, as discussed in Chapters 2 and 3, and local oscillator noise which can be minimized by choosing a high i.f. as explained in Chapter 4.

Owing to the development of the balanced mixers described in §4·5, local oscillator noise is no longer such an overriding consideration as it was previously.

The objection to a high intermediate frequency on grounds of noise factor has also been reduced in importance to some extent, due to development of neutralized-triode circuits, as explained in Chapter 3.

Automatic-frequency correction (a.f.c.) is usual in microwave radar receivers, and if the 'image' tuning point is within the automatic-control range there is a risk of the a.f.c. locking just off-tune to it (see Chapter 8, p. 140); this defect can often be avoided by using a sufficiently high i.f., the alternative being considerable elaboration of the a.f.c. circuit. With a low i.f., a.f.c. difficulties are further increased as a consequence of rectification of the transmitter r.f. pulse in the mixer; the d.c. pulse so produced is usually steep-sided enough to have components in the i.f. range which are amplified in the usual way but are not dependent on the local oscillator tuning. As these affect both halves of the discriminator circuit equally, they can, if small, be balanced out, leaving the operation more or less normal. If large, they use up a substantial part of the grid base of the amplifying valves, the wanted output obtainable is reduced, and performance suffers even if the more critical balancing adjustment is acceptable.

In microwave receivers, coupling between the local oscillator and r.f. circuits is usually reduced to a small value by reactive attenuation in the feed from local oscillator to mixer, as explained in Chapter 4, p. 65. In some lower frequency mixers, however, there is a tendency for pulling of the local oscillator frequency by the r.f. circuits, and this may have to be avoided by choosing a sufficiently high value of i.f.

Most of the foregoing considerations appear to favour the use of as high an i.f. as possible, but an important and frequently overriding argument in favour of a low i.f. is provided by the necessity of maintaining the required amplifier characteristics under all conditions of use and even misuse. Since the tuning capacity in wide-band amplifiers is mainly provided by the valves themselves, it is subject to much the same tolerances as the valve interelectrode capacities and may change if a valve is replaced; such changes should not necessitate re-alinement, and, therefore, the ratio of i.f. to band-width must be sufficiently low, an upper limit being set by valve tolerances and by the amount of mis-alinement which can be permitted. By going perhaps a little lower in i.f. than this limit, an amplifier can be designed with no adjustments, the coils being wound to standard values of inductance before assembly; this is advantageous from the manufacturing point of view, as well as obviating subsequent mis-alinement in service such as that due, for example, to unjustified tampering with trimmers. A ratio of the order of four to one appears to be generally satisfactory for this purpose in the case of wide-band amplifiers.

Chapter 7

TRENDS IN PRACTICAL RECEIVER DESIGN

The noise-factor concept and the wide-band amplifier development discussed in previous chapters have derived their main stimulus from radar, but are applicable—actually or potentially, and in varying degrees—to many other kinds of receiver. In turning our attention to receivers for broadcasting or communications, however, the main interest is to be found in the general trends of engineering development, and reduction of noise factor is only a relatively small part of the story.

The following attempt at assessing tendencies in receiver design has been restricted to broadcast, communication and television receivers, partly because these classes of receiver are the most important numerically and afford the largest amount of material for such a purpose, and also because it is hardly feasible to deal adequately with the more specialized types of receiver without describing the techniques such as direction-finding, pulse communication, or radar to which they belong.

7·1. Broadcast receivers

7·1·1. *General*

For obvious reasons the war years have not been conducive to advances in broadcast reception, but trends which were beginning to appear before the war have now been further stimulated by the export drive. This applies particularly to the improvement of short-wave performance.

As a background for discussion it is useful to define the ideal broadcast receiver, and it will probably be agreed that this would be one capable of receiving enjoyable music or intelligible speech from any broadcasting station except in so far as this is prevented by fundamental limitations, and conditions external to the receiver. The reproduction must be as pleasing as possible, subject to the same limitations, and the best possible results should be readily obtainable by any listener of normal intelligence. This ideal is, of course, largely subjective, and cannot be put into precise engineering terms; for example, distortion unnoticed by one listener can ruin the enjoyment of another. It is interesting to examine a 'most typical' modern broadcast receiver in the light of this definition; such a receiver is probably a table model comprising a triode-hexode frequency changer, variable-mu i.f. stage, diode-triode for detection, a.v.c. and l.f. amplification, and a high slope output pentode. The band-width is about 8 kc./sec. to 3 db. down, and the selectivity with respect to adjacent channels is about 30 db., which is adequate to remove direct modulation interference, with the standard 9 or 10 kc./sec. frequency separation, except when the interfering signal is so strong that reception is ruined in any case by 'side-band splash'. With a reasonably good aerial, sensitivity is usually limited on medium or long waves by external noise or interference, and we are left with at most two major deficiencies in the receiver; the first is failure to reproduce all the available side-band frequencies, and the second is the general need for some measure of skill in tuning to ensure satisfactory reproduction.

The latter deficiency can be largely overcome by accurate scale calibrations, and by such devices as the 'Magic Eye', push-button tuning and automatic-frequency correction, provided these are properly engineered, and in this connexion there are a number of points easily overlooked; a.f.c. systems are usually capable of locking several kc./sec. off-tune, and modulation should therefore be suppressed when the error in manual tuning exceeds the amount that can be corrected adequately; push-button systems without a.f.c. should have a maximum drift not exceeding 1 or 2 kc./sec., and with a.f.c. the drift must be less than half a channel width, i.e. less than 4½ kc./sec., in order to ensure that the right station is obtained; tuning indicators should be operated via a sufficiently selective circuit not included in the a.g.c. loop.

Progress in long- and medium-wave tuning systems suffers from the continual drive to cut production costs and perhaps insufficient education of the public. Automatic-frequency control is not yet in very general use, but push-button tuning has a strong popular appeal in spite of the deficiencies of some systems, and there is a well-defined trend in the direction of improved reliability; this has involved the development of high stability circuit elements for oscillator tuning. One satisfactory system employs a specially designed tuning condenser with accurate mechanical location in present positions, and another uses separate pre-set inductances for each station, tuned by fixed silver-on-mica condensers. Fixed inductances have also been used successfully in conjunction with ceramic-type trimmers. Motor-driven systems in general require the addition of a.f.c. It is on short waves, however, that the tuning problem is most urgent as well as most difficult to solve, and it is being tackled with a large measure of success as described in § 7·1·2.

Quality of reproduction is limited at the lower end of the register primarily by cabinet size, and at the upper end by adjacent channel interference, and the quality of transmissions. Under favourable conditions it is possible to obtain a substantially higher order of realism than can be expected from the typical 'table-model' receiver, but cost is higher, and a receiver designed for this purpose requires considerable elaboration and skill of operation if, under more usual conditions, it is even to be the equal of the table model, because any unwanted content in a received signal—such as heterodyne whistles, background noise, or high-order harmonic distortion—tends to be much more objectionable than the absence of wanted constituents. Even at 5 miles from a powerful broadcasting station, the heterodyne whistle from an adjacent channel station hundreds of miles away can, in the author's experience, render 'high-fidelity' reception intolerable after dark unless a 'whistle filter' is fitted, which involves loss of *some* of the wanted side-bands, and (since it should be optional) an additional control for the listener to manage.

In view of these difficulties it is not surprising to find there are no very marked tendencies in the direction of better reproduction. The more expensive table models incorporate variable selectivity and whistle filters, and are capable of reproducing side-band frequencies up to 7 or 8 kc./sec., and 'Console' type receivers provide improved bass response, but with the cheaper class of table set there is little to choose as regards quality of reproduction between current models and the better examples of say 15 years ago, and although we *can* obtain better quality than this by paying for it, there is not very far to go before we find the cost increasing rapidly for ever smaller returns. This situation will remain until F.M. or other forms of ultra-short-wave broadcasting become properly established. It will be interesting to await developments in this direction, and perhaps one day we may be able to enjoy the realism associated with binaural reception.

The main obstacle to 'local-station' quality of reproduction in the case of distant stations on medium and short waves is selective fading. The possibility of some improvement by means of carrier boosting is discussed in Chapter 9, but it remains to be shown whether the methods suggested can be successfully employed for broadcast reception.

It is in the matter of short-wave reception that most scope can be found for improvement; this is of considerable importance for export and is therefore likely to receive increasing attention. The need to simplify short-wave tuning has already been mentioned, but this is far from being the only problem. Let us see what happens when we try and tune our 'typical' broadcast receiver to a short-wave station which is just capable of giving worth-while reception on, say, a good communication receiver; assuming we have mastered the technique of finding the station and tuning it for optimum reception, we may be lucky and obtain just as good a result as in the case of the communication receiver, but in all probability the station will be swamped by image interference or receiver noise, or the gain will be inadequate to cope with fading. If we *are* able to listen to the station, before long we probably find that fading and distortion increases, due to tuning drift, and rather than face repeated uprisings from the armchair we lose patience and switch off, or perhaps over—permanently—to medium waves. It is unfortunately true that short-wave reception can hardly ever be as good as reception from the local station, and usually justifies itself, at best, by a relatively narrow margin depending on the individual listener but leaving little or nothing to spare for imperfections in the receiver. The defects listed above are removable as described in § 7·1·2, by known techniques, but not without an increase of some 20–40% in cost. Whether half-measures are worth while or not is a matter of opinion, and although much effort is being put into easier tuning it would appear that so far only a small number of manufacturers are impressed with the importance of image interference and tuning drift.

Other interesting developments include a new constructional technique aimed at reducing production costs discussed briefly in § 7·1·3, and progress in the direction of the personal portable receiver based on the miniature valves and components developed during the war. A typical receiver with built in batteries and L.S. weighs 4 or 5 lb. and occupies a volume of about 100 cu. in., and further reduction of size and weight could presumably be effected by using say a deaf-aid earphone in place of the L.S. A possible function of such receivers is to enable listening to news bulletins or special programme items in the bus or train, when walking or cycling, etc.—but in making a receiver as small as possible compromises must be made in regard to battery size, etc., and the result may prove uneconomical for sustained listening.

7·1·2. *Improvements in short-wave reception*

7·1·2·1. *Signal-to-noise ratio.* The conversion conductance of a typical frequency changer, say a triode-hexode, is only about one-tenth of the slope of an EF 50, although the anode current is about the same. The proportion of partition noise is of the same order, or worse, than in the EF 50 case, so that referring to § 3·9 we can make a fairly good guess at the noise resistance in spite of the scarcity of official data, and it comes to something of the order of 100,000 ohms. Typical values for the input circuit impedance (R_c) with no aerial loading are respectively 4000 and 15,000 ohms at the longest and shortest wave-lengths in the 'short-wave' range. If we refer to equation (3·3), we find that the noise factor under these conditions becomes very approximately $4R_n/R_c$, or 14 to 20 db. With some frequency changers, e.g. pentagrids, the noise resistance may be much higher, and it is also to be noted that the optimum input coupling (which approximates to a power match), as well as aerial efficiency, is very dependent on frequency. With one aerial system, at a given frequency, the range limitation may be determined by external noise, and a receiver with a noise factor of 20 db. may give results indistinguishable from those obtained with 2 db. noise factor. Another aerial system, either less efficient or less well suited to the particular input coupling (which is usually fixed), may be perhaps 20 db. down in performance; the results with the 2 db. receiver will not be much affected, but the 20 db. receiver will probably be quite useless. It is usual, therefore, in the better class of receiver, to precede the frequency changer with an r.f. stage. For lowest noise factor this should be one of the high-slope pentodes such as the EF54, EF50 or SP41, and such receivers first made their appearance on the market a year or two before the outbreak of war. The r.f. stage, especially with a high slope valve, is an embarrassment on medium and long waves, since it is usually unnecessary, and likely to accentuate the various overloading and cross-modulation effects produced by strong signals; in one such receiver, a medium and long-wave inverse feed-back control, pre-set to suit local conditions, was provided on the r.f. stage, and in another the r.f. stage was cut out entirely at these wave-lengths.

7·1·2·2. *Image rejection.* Provision of an r.f. stage helps to some extent with image rejection, owing to the additional tuned circuit which goes with it, but the improvement is small compared with the desirable 40 db. or so unless the i.f. is of the order of a megacycle or more. The inclusion of s.w. ranges was the principal factor in causing many manufacturers to change over from an i.f. of 110–120 kc./sec. to one of 450–475 kc./sec., but this is about as far as it is possible to go without getting into difficulties with the medium and long wave-bands. One way out of the difficulty is to use double-frequency changing (d.f.c.), and another is some form of image cancellation.

Special problems involved in d.f.c. are the additional images and harmonic responses. The additional images arise as follows:

Let f_1 be the first i.f. and f_2 the second; the second oscillator frequency will then be $f_1 \pm f_2$, and a signal of $f_1 \pm 2f_2$, if allowed to reach the second frequency changer, would cause an image response. If f_w is the frequency of the wanted station the first oscillator will have a frequency $f_w \pm f_1$. Taking (by way of example) plus signs, there is an image frequency which will be $f_w + 2f_1$, and, in addition, signals at $f_w - 2f_2$ or $f_w + 2f_2$ will beat with the first oscillator to give the frequency $f_1 \pm 2f_2$. Fortunately, it is fairly easy to prevent these signals from reaching the second frequency changer, by means of circuits selective to f_1; otherwise three image responses will be found.

Harmonics of the second oscillator present a worse problem. Some voltage at each harmonic is almost certain to find its way on to the grid of the first frequency changer, most probably without passing through the pre-selector circuits so that there is no image rejection applied to it and it breaks through at two points in the tuning range. Two ways of tackling this problem have been successfully employed in broadcast receivers; by using the lowest value of second i.f. consistent with obtaining the desired image rejection the harmonics are relatively weak, being high-order, and can be rendered negligible by careful lay-out and the use of as much gain as possible in front of the first mixer. Also if the second local oscillator is used for band-spreading, 'whistles' due to the nth harmonic are n times as sharply tuned as wanted signals and easily avoided by a small tuning shift. In one such receiver the i.f.'s were 1 Mc./sec. and 120 kc./sec. respectively. The other alternative requires the ***highest*** i.f., consistent with rejection of $f_1 \pm 2f_2$ and care in choice of the exact frequency, so that the relatively few harmonics and their images fall outside the broadcast bands. In one receiver produced shortly before the war a first i.f. of 3·1 Mc./sec., a second i.f. of 465 kc./sec., and a second local oscillator frequency of 2·635 Mc./sec., were employed successfully.

The idea of image cancellation, common in medium-wave broadcast receivers having a low i.f., is attractive as an alternative to double mixing, but not very easy to apply in its usual form because of the low ratio of the i.f. to the mean frequency and the wide range to be covered, and because the usual methods require a pair of coupled circuits. However, satisfactory image rejection can be achieved with a single circuit, and fig. 7·1 shows one arrangement which has been used successfully. This circuit, used as an intervalve coupling impedance, would give maximum gain at the frequency $1/2\pi\sqrt{(LC)}$. At any lower frequency, which can be the image frequency, the circuit LC is inductive, and a series resonance, giving a very low impedance, can be obtained between A and B by suitable choice of C_s. Fig. 7·2 illustrates the practical application of this device, the operation being somewhat modified

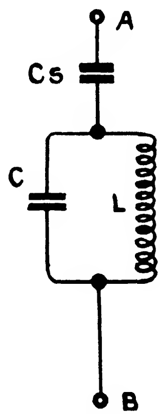


Fig. 7·1. Basic circuit of the simple image rejector described in the text. In the absence of circuit losses, the impedance between A and B would be infinite for the wanted signal and zero for the image.

by the capacity C_p which is made up of interelectrode and other stray capacities. This has the effect of reducing the gain by the factor $C_s^2/(C_s+C_p)^2$, besides modifying the values required for C and C_s. The gain reduction means that to achieve adequate gain at the higher frequency s.w. broadcast bands, the

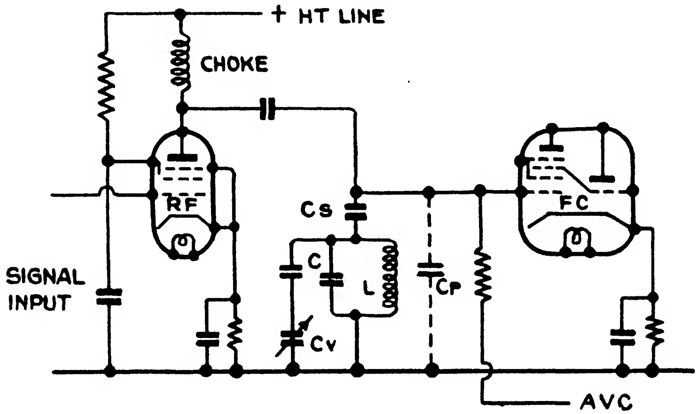


Fig. 7·2. Showing application of image rejector to a practical design: switching complications omitted. C_p represents the stray capacities across the coupling impedance. For band-spreading a section C_v of the gang condenser is connected, in series with a small capacity, across the circuit LC.

r.f. stage *must* employ a high-slope valve, but as previously explained this is also desirable from considerations of noise factor.

The conditions for series and parallel resonance (i.e. image rejection and signal acceptance) are given respectively by

$$f_i = 1/2\pi\sqrt{(LC')} \quad \text{and} \quad f_s = 1/2\pi\sqrt{(LC'')},$$

where $$C' = C + C_s, \quad C'' = \frac{C_p C_s + C_p C + C_s C}{C_p + C_s}. \qquad (7·1)$$

To simplify the algebra C has been taken as the total capacity across L.

If r is the ratio of intermediate frequency to f_1, the required value of C_s is given by

$$C_s \simeq 2r(C_p + C) + \sqrt{[4r^2(C_p+C)^2 + 4rC_pC]}. \qquad (7·2)$$

If Q is the ratio $\omega L/R$, the image rejection ratio is given approximately by

$$Q^2(1-\omega^2 LC)^2 C_s^2/(C_s+C_p)^2. \qquad (7·3)$$

Fig. 7·3 shows a typical response curve for the image-rejector circuit. It can be inferred from this that there is reasonable latitude of design to allow for component tolerances, and also for band-spreading by the methods advocated in the next section, although the optimum value of C_s depends on frequency. The amount of image rejection, and therefore the allowable degree

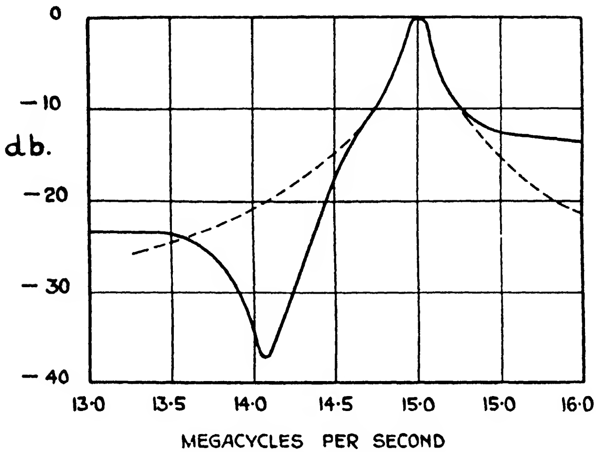


Fig. 7·3. Measured selectivity of image rejector at 15 Mc/sec. The estimated curve for a single tuned circuit with similar losses is shown in dotted lines. In practice some additional selectivity is given by the aerial circuit.

of departure from strict adjustment, increases rapidly with wave-length.

This circuit is particularly suitable for a band-spread system employing switched inductances as described later, but could no doubt be applied with condenser tuning by the addition of an auxiliary variable capacity to perform the function of C_s.

7·1·2·3. *Short-wave band-spreading and reduction of frequency drift.* Most short-wave broadcasting stations are to be found within seven narrow bands all less than 300 kc./sec. in width and centred around 6·1, 7·25, 9·6, 11·8, 15·22, 17·8, and 21·6 Mc./sec., and yet it is common practice to cover the entire frequency range, or at least up to 18 Mc./sec., by capacity variation using a single set of tuning inductances. As each station's share of the tuning scale is less than 1/1000th part, it is not surprising that tuning is apt to be a highly skilled operation. Band-spreading is therefore receiving increasing attention, its object being to spread out the wanted bands so that they occupy a length of tuning scale long enough to be read to within a few kc./sec., and at the same time to make it easy to set the tuning to within, say, the nearest kilocycle. There are many ways of doing this, as illustrated in fig. 7·4. The left-hand side of this 'family tree' embraces a very large number of ingenious mechanical devices, but these suffer in general from the defect of backlash, and it is doubtful if there is a complete solution of the problem suitable for mass production, although very successful precision drives with geared-up scales have been employed in communication receivers. With a high degree of mechanical band-spreading, a device such as a flywheel is needed to enable rapid tuning from one band to another.

At this point we have to return to the theme of frequency drift. Any mains-operated receiver which attempts to cover more than one, or at most two, of the short-wave broadcasting bands by capacity variation using a single tuning inductance must almost inevitably suffer from this defect. Let us consider a typical case, namely, a 3 to 1 wave-length coverage involving a 9 to 1 capacity change; the minimum capacity which can be achieved is usually in the region of 50 pF., making the maximum about 450 pF. (which incidentally means a loss of several decibels in noise factor owing to the low circuit impedance, a fault which has to be accepted). Frequency drift is mainly due in practice to capacity changes during warming up—valve capacity and coil self-capacity in particular—so that as a percentage it increases as the square of the frequency, and in kc./sec. as the cube of the frequency, being usually of the order of 40–80 kc./sec. on the 16 m. band. The valve, which is the main offender, warms up relatively quickly, but drift is liable to be serious for anything up to an hour after switching on. Small compensating capacities, operated, for example, by a bimetal strip, have been employed to neutralize drift, but are only a partial solution to the problem owing to the different heating and cooling rates of the various components.

By using a separate set of inductances for each of the broadcasting bands, selected by a switch or push-buttons, both the band-spreading and frequency-drift problems can be solved simply and efficiently. Local oscillator valve and coil capacity changes may then be swamped by using fixed tuning capacities of the largest values compatible with adequate efficiency, and the drift of other components may be compensated by letting the capacity consist of at least two parts suitably proportioned and having positive and negative temperature coefficients respectively. Band-spreading may be applied in various ways, as indicated in the

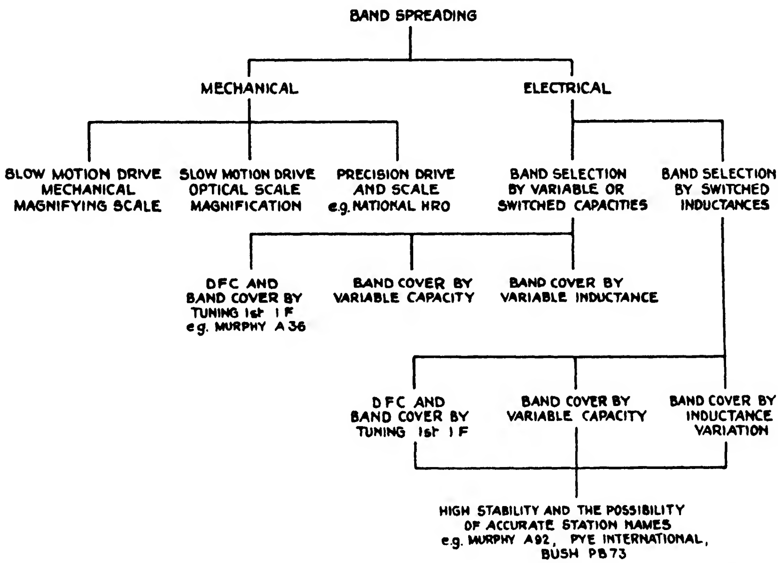


Fig. 7·4. 'Family tree' depicting the various methods of band-spreading.

'family-tree' (fig. 7·4). In one such receiver the normal medium- and long-wave tuning condenser is used with a series capacity, as illustrated in fig. 7·5, which produces a band-spread tuning curve of the form shown in fig. 7·6. This does not look ideal at first sight, but the total frequency coverage is proportional to the mean frequency, and a more or less constant degree of band-spreading is achieved by using the region *AB* for the higher frequency and *CD* for the lower frequency bands. This receiver incorporates the image-rejection circuit described above. An even better method of band-spreading is possible in conjunction with double-frequency changing, since by varying the second local oscillator frequency a constant spread is obtained on all bands and a very accurate 'relative' calibration is obtainable. This system was employed on both of the double-mixing receivers mentioned earlier, but does not so far appear to have been used in conjunction with separate inductances for each band. One problem confronting the designer of band-spread systems is whether to spread the r.f. circuits as well as the local oscillator, or leave them pre-set to the middle of

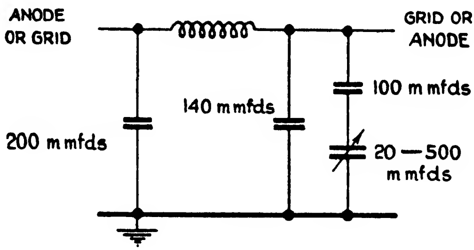


Fig. 7·5. Arrangement of Colpitt's oscillator for s.w. band-spreading using the main tuning condenser.

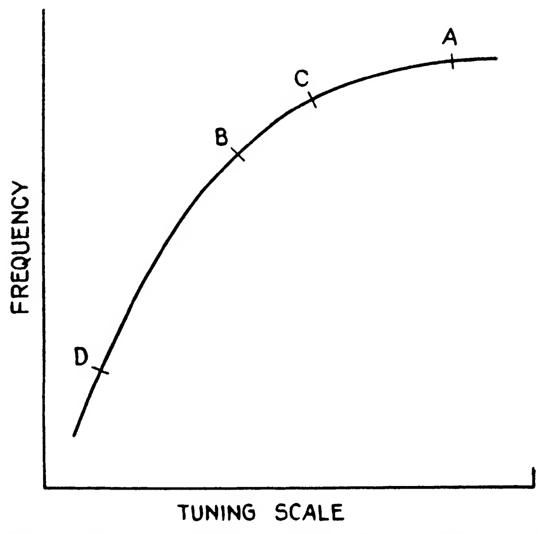


Fig. 7·6. Scale calibration obtained with the arrangement of Fig. 5. The portion *A–B* is used for the higher and *C–D* for the lower frequency bands.

each band. In general, the input circuit is much flattened, and its precise tuning point rendered rather indefinite, by the aerial impedance which is usually an unknown quantity; 'spreading' is therefore rather pointless. The slight loss of gain and noise factor at the edges of the normal s.w. broadcast bands which results from not varying the tuning of the r.f. intervalve circuits is usually acceptable as such, but tuning has been found advisable in the case of the image-rejector circuit.

Band-spreading makes accurate 'logging' of s.w. stations possible, and consequently in addition to the warming-up drift already discussed, any long-term frequency drift is undesirable. Knowledge of the causes of the changes which take place over long periods is incomplete, but it is known that mica compression-type trimmers are inadvisable for use in the oscillator circuit, and fixed capacities which affect the tuning should be of air-dielectric, silver-on-mica or ceramic construction. Inductances do not usually give much trouble assuming reasonable care in design, including wax impregnation to protect them against humidity. With these precautions it was found possible in the case of the switched-inductance receiver already mentioned to put short-wave station names on the dial with a fairly high degree of probability that the stations would be found as marked, i.e. a calibration accuracy of the order of $\pm 0{\cdot}025$ to $0{\cdot}05\%$ was obtained.

7·1·3. *New methods of radio production*

Many ideas have been put forward for simplifying the mass production of receivers, in which the wiring operation is carried out more or less automatically, the conductors being deposited, for example, on an insulating sheet by spraying metal through a stencil. What is believed to be the first practical application of such methods to domestic radio production has recently been described by J. A. Sargrove,* and in this process the idea is carried considerably further to include the prefabrication of coils, resistors and capacities. Metal is deposited on both sides of plates of insulating material suitably moulded so that after a subsequent surface milling operation we are left with: (*a*) grooves filled with metal, which constitute the connecting wires and coils, the latter being of spiral form, (*b*) depressions, with associated metallizing on both sides of the sheet; the metallizing forms the plates of condensers for which the sheet itself, greatly reduced in thickness at the depressions, provides the dielectric.

Valve-pin contacts are secured to the sheet by means of eyelets. Switch contacts are similarly attached, in conjunction with a

* *J. Brit. Inst. Radio Engrs*, Jan.–Feb. 1947.

suitable rotor disk let into the insulating sheet. Resistors are deposited on the sheet either by printing technique, or by spraying graphite through stencils, the higher wattage resistors being deposited on the walls of the cabinet where a larger surface is available for the purpose of heat dissipation. Multi-plate capacitors are formed by alternate sprayings with metal and lacquer. Tuning capacities are semi-variable, using a spring-like flexible vane or a rotating metallized disk as the moving element.

Manufacture and circuit testing is fully automatic and electronically controlled, apart from such operations as insertion of the movable condenser vanes, plugging in valves and electrolytic capacities, etc.

So far this method has only been applied to simple kinds of receiver, but it is interesting at least as a departure from conventional methods, and may well prove to have far-reaching possibilities.

7·2. Communication receivers

A communication receiver has recently been defined* as one 'not designed for limited or specific purposes', thus automatically excluding receivers for broadcasting, television, radar, etc. This seems to be a good starting point, and we can proceed, as in the case of broadcast receivers, with an attempt to define the ideal. Such a receiver would presumably be one capable of receiving any transmission intended for the communication of any form of information from one point to another, and able under the prevailing conditions, external to the receiver, to provide intelligence; in other words failure to receive any signal, or any defect in reception, must be attributable entirely to circumstances external to the receiver.

Many of the considerations already discussed in § 7·1 apply equally to communication receivers, but additional problems are presented by the wide frequency coverage demanded, which should ideally embrace all transmissions except radar and a few other specialized types, and by the very narrow band-widths required for optimum reception of Morse signals. In practice it is necessary to stop far short of our supposed ideal, particularly as

* G. L. Grisdale and R. B. Armstrong, 'Tendencies in the design of the communication type of receiver', *J. Instn Elect. Engrs*, 93, part III, p. 365.

regards frequency coverage, otherwise the receiver would become so complicated and expensive that on balance its range of usefulness would be severely restricted instead of extended. The frequency coverage may be as great as 60 kc./sec. to 30 Mc./sec., but is usually somewhat less, 2 Mc./sec. being a typical lower limit. Some receivers cover up to 60 Mc./sec., and others are designed to cover ranges such as 30–300 Mc./sec. Above 30 or 60 Mc./sec. the applications tend to become more specialized and design requirements are considerably modified, so that it seems reasonable to follow the precedent of Grisdale and Armstrong (loc. cit.) and take an upper limit in the region of 30 Mc./sec.

7·2·1. *Band-spreading and frequency drift*

The narrow band-width (sometimes as low as 100 c./sec.) required for c.w. reception makes it essential to tackle the problems of band-spreading and frequency drift very thoroughly, and at the same time the requirement of a wide and continuous frequency coverage makes it extremely difficult to do so. In typical receivers the coverage of any one band is therefore restricted to an octave or less. We can employ any of the electrical methods of band-spreading discussed in § 7·1, or alternatively slow-motion drives with precision gearing and 'logging scales' geared up from the main tuning condenser spindle and readable to whatever degree of accuracy is justified by the stability of the circuits. Such slow-motion drives are usually provided with some form of 'fly-wheel' action to enable rapid progression from one end of the band to another. Use has been made in some communication receivers of double mixing, with band-spreading by means of the second local oscillator as described in § 7·1·2·3.

Limitation of the tuning range to an octave is helpful from the point of view of frequency drift, since the minimum oscillator capacity can be greater or the valve can be tapped down into the circuit. Even so, and with the highest quality components, drift remains a serious problem,* and as a further measure bimetal strip compensating capacities have been employed successfully. Negative temperature-coefficient ceramic capacities as mentioned in § 7·1·2·3 have also been used.

* Grisdale and Armstrong (loc. cit.) quote as typical a figure for the frequency temperature coefficient of 50 parts in 10^6 per °C.

7·2·2. *Noise factor*

Estimates of the frequency below which external rather than internal noise is usually the limiting factor vary from about 2 to 15 Mc./sec. depending largely on the aerial system. The reduction of noise factor has received considerable attention, and noise factors of less than 2 db. have been obtained up to 16 Mc./sec., but at 30 Mc./sec. the performance is usually rather poor, say 8–16 db. It should be possible to improve considerably on this latter figure, using the principles of design outlined in Chapter 4.

7·2·3. *Selectivity*

For c.w. reception a communication receiver is usually only required to handle perhaps 5 or 10 impulses a second. It must do so without appreciable distortion or 'slurring of morse characters', but a band-width of 100 c./sec. should usually be more than adequate from this point of view. However, to use such a band-width at a frequency of 15 Mc./sec. requires a short-period local oscillator stability of 2 or 3 parts per million, and it is difficult to guarantee this even with temperature compensation. The time required to find a transmission for which the dial setting is not accurately known is inversely proportional to the band-width; telephony, of course, requires a band-width of several kc./sec., and certain other transmissions, due either to their general character or insufficient stability, may also need a wider band-width than the minimum. In view of these various considerations, variable selectivity is essential unless the receiver is restricted to certain kinds of communication only, or unless large departures from the optimum performance are tolerated. It is usual also to include optional low-frequency selectivity with a band-width of say 100 c./sec. at a mean frequency of 1000 c./sec., to give additional adjacent channel selectivity in the case of c.w. This does not greatly diminish the need for i.f. selectivity, important functions of which include prevention of the overloading of later i.f. stages by strong unwanted signals, and provision of the 'single-signal' facility; this is the rejection of unwanted signals producing the same beat note as the wanted signals but situated on the opposite side of the beat oscillator-frequency, and is analogous to the use of preselection to keep r.f. images out of the i.f. amplifier.

Any desired band-width, down to about 1½ or 2 kc./sec., is readily obtainable by switched couplings in i.f. amplifiers, and band-widths from about 100 c./sec. up to say 1 or 2 kc./sec. can be achieved with simple forms of crystal filter.

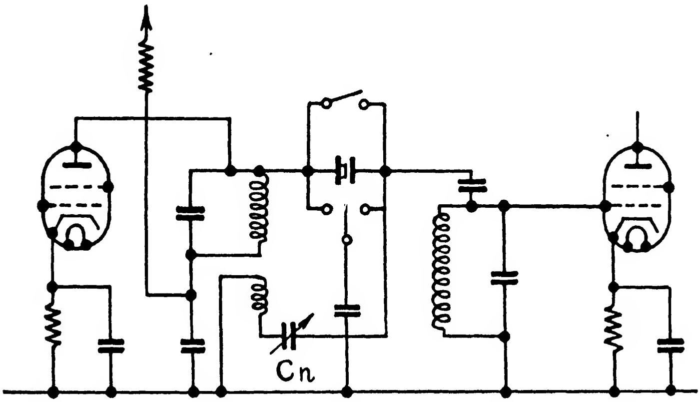


Fig. 7·7 (*a*). Typical crystal filter. The switched capacity enables the band-width to be altered by staggering the tuning of the primary and secondary circuits.

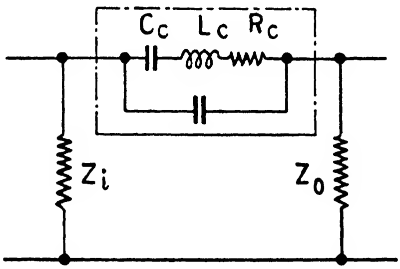


Fig. 7·7 (*b*). Equivalent circuit to illustrate action of crystal filter.

Fig. 7·7 illustrates a typical crystal filter.* The capacity C_n neutralizes the capacity between the plates of the crystal-holder, and the crystal itself behaves as a series resonant circuit C_c, L_c, R_c. The effective band-width is adjustable subject to practical limitations, by varying the impedances Z_i, Z_o of the input and output circuits, either by using switched resistances or, as shown, by detuning. Typical values are $Z_i = Z_o = 75{,}000$, $R_c = 5000$,

* Grisdale and Armstrong, loc. cit.

$L_c = 25$H, i.f. = 465 kc./sec., band-width = 1710 c./sec., and (with 7·6 pF. detuning of the side circuits) 500 c./sec., the gain being roughly constant at 37 db. for a valve with a slope of 2mA/volt.

7·2·4. *Automatic gain-control*

The provision of a.g.c. is usual, but the optimum design of a.g.c. circuits presents some difficulty as the requirements vary greatly, depending on the nature of the signals and on the nature of fading. As discussed elsewhere (Chapter 9), an a.g.c. system which holds the output level constant to close limits over a wide range of carrier level is an embarrassment for telephony reception in the presence of selective fading; on the other hand, most designers of communication receivers seem to consider such a system desirable. A long 'off' time constant is necessary for c.w. reception in order to maintain a constant noise level during the spaces. An a.g.c. on/off switch is usually fitted.

7·2·5. *Noise limiters*

Interference in s.w. reception is often of the 'impulse' type, such as that due to motor-car ignition systems. Such impulses are usually of relatively short duration and only cause serious trouble because their amplitude is very much larger than that of the wanted signals. This means that they can produce various forms of paralysis and shock excitation. The requirement is therefore one of preventing the signal amplitude at any point in the receiver where trouble can arise from exceeding, say, the equivalent of 100% modulation of the wanted signal.

The output circuit of the second detector is usually a suitable place for the introduction of amplitude limiting, which can be carried out simply as follows. A condenser is charged up to the mean carrier voltage, the time constant being fairly long. This voltage is applied as bias to a diode or diodes arranged in series or parallel with the signal output lead in such a way that if the instantaneous signal amplitude tries to exceed the bias voltage the diodes either conduct or are cut-off, and constitute open or short circuits, whichever is appropriate. The limiting level corresponds to 100% modulation if the output feed is taken from a point such that the mean level is equal to half the bias voltage.

Such an arrangement is shown in fig. 7·8;* the bias is developed across the 0·05 μF. condenser, and for modulation percentages up to 100, D_3 conducts and D_2 is an open circuit. For the duration of interference corresponding to more than 100% modulation, D_3 tends to become an open circuit, and any excess voltage which

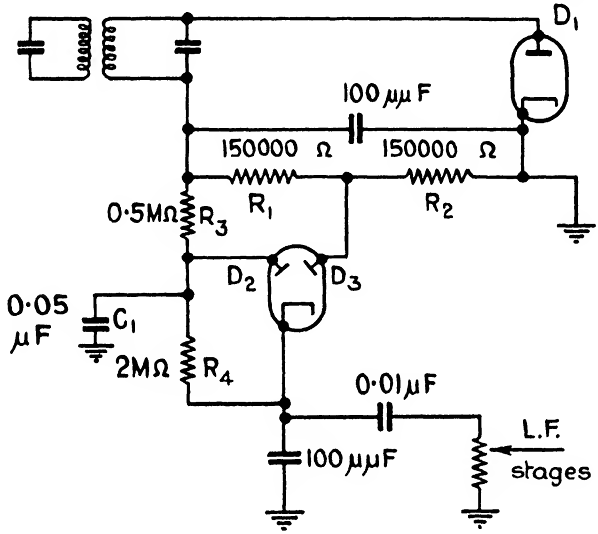


Fig. 7·8. Typical noise-suppression circuit.

may nevertheless get past D_3 (due, for example, to leakage through the capacity of the diode) is clipped off by D_2 which becomes conducting. Either D_2 or D_3 is sometimes omitted.

7·2·6. '*Specialized*' *communication receivers*

The definition of the term 'communication receiver' which we have been using hitherto is open to criticism in so far as many receivers used for communication have some specific job to do, and improved performance or lower cost (or both) is achieved by limiting them to this purpose.

The 'communication receiver' came into being in the first instance in response to amateur needs, and most of the available receivers seem intended to comply as far as possible with the definition on p. 120. However, since amateurs are restricted to a limited number of narrow bands, such wide coverage appears to

* Grisdale and Armstrong, loc. cit.

be unnecessary and improved performance—particularly in the matter of frequency drift, as already explained, but also as regards noise factor at the higher frequencies—could be obtained by limiting the tuning ranges so that they are no greater than is really necessary.

It sometimes happens that a receiver is intended for receiving some specified transmitter or transmitters, and in such cases the tuning and frequency drift problems have been solved by the use of crystal control of the receiver as well as the transmitter. By using various combinations of crystals, and their harmonics, a fairly wide choice of operating wave-lengths can be arranged.

Some receivers, particularly communication receivers for shipboard use, may be required to work in close proximity to powerful transmitters, and cross-modulation becomes a serious problem. In such cases it is often necessary to sacrifice noise factor in the interest of input circuit selectivity. In § 9·5 devices are described which have recently been introduced for reducing cross-modulation without elaborate filtering or pre-selection.

7·2·7. *Future possibilities*

Some receivers already incorporate built-in frequency calibrators, and an extension of built-in testing facilities can reasonably be expected. Consideration of the performance specifications of communication receivers shows that there is still much scope for improvement, particularly as regards noise factor at the highest frequencies where it is most needed, and the maintenance of such an improvement would be greatly facilitated by the use of built-in 'noise diodes' of the type described in Chapter 5, which are extremely simple, accurate and easily operated.

Reduction of size, by the use of miniature valves and components, and improved 'proofing' against climatic conditions, are other directions in which further developments can be expected on the basis of recently acquired knowledge and experience.

7·3. Television receivers

7·3·1. *General*

Television receivers comprise a straight or superheterodyne portion followed by a detector, a video amplifier feeding signals to the grid of the vision tube, synchronizing and time-base circuits, and a sound channel. The wide-band width required, and the fact that for some time to come receivers in this country will not require to be tunable to more than one station, makes the use of a 'straight' receiver a practicable proposition. The superheterodyne, on the other hand, offers certain advantages such as reduced risk of instability, less likelihood of the intervalve coupling circuits getting out of adjustment, and easier sound-channel rejection. Time bases tend to follow pre-war practice, but the Miller-transitron,* originally developed for radar, is becoming popular; this uses a single pentode to produce a linear saw-tooth.

The sound and vision channels usually share the first few stages of the receiver, separating after the mixer in the case of a superheterodyne. Sound-channel rejecting circuits are frequently employed in the vision channel and are normally necessary for double side-band reception if the full vision band-width (6 Mc./sec.) is employed, because the sound channel is only 0·5 Mc./sec. away from the edge of the vision band. It is not essential, however, to use both side-bands, and some receivers are alined so that the carrier is about 6 db. down on the l.f. side of the pass-band; for low modulation frequencies reception is double side-band, and one of the h.f. side-bands is doubled in amplitude relative to the carrier so that elimination of the other side-band does not entail a loss of h.f. response. The tuning and alinement of such a system is of course somewhat more critical than for the normal case.

To minimize the effects of ignition interference amplitude limiters are usually fitted in both sound and vision channels and many ingenious circuits are in use. The use of a wide i.f. band for the sound channel assists the operation of noise limiters by ensuring rapid decay of the i.f. oscillations set up by the interference.

7·3·2. *Frame synchronizing*

The frame synchronizing signal is usually derived by integration of the pulses, and the time-base is fired when a sufficient amplitude has been built up. The time at which this critical amplitude is reached is liable to be affected by various forms of disturbance including interference, mains variations, fading, etc., and any change taking place in a time comparable with the frame period will obviously tend to upset the interlacing.

* See, for example, W. T. Cocking, 'Linear saw-tooth oscillator', *Wireless World*, June 1946, for a detailed account of this circuit.

Although satisfactory results *can* be achieved, the alternative method* of frame synchronizing, illustrated in principle by fig. 7·9, is of particular interest as it strikes at the root of the difficulty. The synchronizing signals are applied to a differentiating circuit of time constant comparable with the frame-pulse duration, and the time-base is triggered by the first of the negative 'pips', i.e. coincidently with the trailing edge of the first frame pulse, so that the timing is accurately determined.

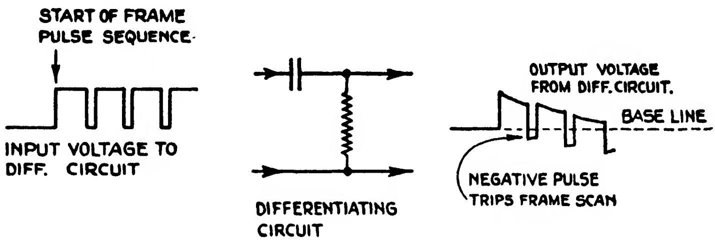


Fig. 7·9. Diagram illustrating principle of differentiation method of frame synchronizing.

7·3·3. *The supply of E.H.T. voltage*

A problem which has received a good deal of attention concerns the production of the high voltage required for operating the cathode-ray tube. High voltage windings on 50-cycle mains transformers require, on account of mechanical strains, to be wound with much thicker wire than that dictated by considerations of current-carrying capacity. This leads to either a rather expensive transformer or a compromise design of doubtful reliability. Another objection to this method of obtaining the E.H.T. voltage is the danger of shock which exists during servicing, in consequence of the size of the smoothing condensers necessary when working with a 50 c./sec. input.

One answer to this problem is to make use of the high voltage developed when a sufficiently large current change is produced suddenly in an inductive load. This is a situation which normally exists in the line-scan output transformer during the fly-back, and although the voltage produced is normally only of the order of half that required to operate the display tube it is possible to obtain sufficient voltage, with the aid of careful transformer

* W. T. Cocking, 'Interlacing', *Wireless World*, April 1947.

design, by using either a step-up winding or a voltage-doubling rectifier. In one method* the line scan is fed to a separate valve having a ringing choke in its anode circuit. With either of these methods, since the 'supply' frequency is 10,000 c./sec., the smoothing capacitances can be small, of the order of 0·001 μF.

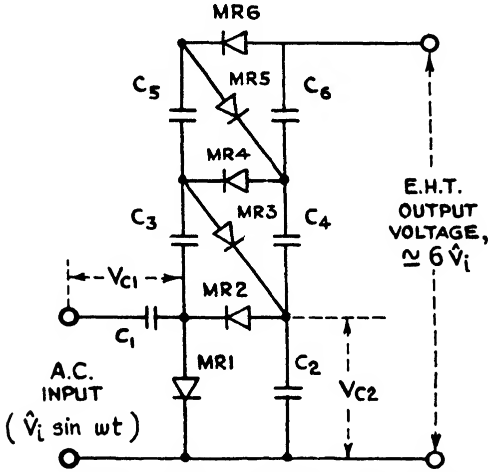


Fig. 7·10. Simple Cockcroft-Walton voltage multiplier.

Another system† employs an r.f. oscillator working at say 20 kc./sec. The coil assembly includes a resonant secondary winding of the highest possible dynamic impedance and feeding a low-capacity high-voltage rectifier. The secondary is coupled so that it constitutes a matched load absorbing all the r.f. power available, and since its impedance, even with the loading of the rectifier circuit, is several megohms the voltage developed is relatively large. As an example, a secondary coil wound with 1800 turns of 41 enamelled and single silk covered wire wound on a miniature four-section r.f. choke former gave a rectified output of 2000 volts 0·5 mA. when over-coupled to an EF 50 oscillator working at 500 kc./sec. Higher efficiencies can be obtained with larger coils working at lower frequencies. The rectifier filament can, if desired, be fed from a low-voltage winding on the r.f. coil assembly.

* C. H. Banthorpe, 'E.H.T. supply for television receivers', *Electronic Engng*, Aug. 1947.

† R. D. Boadle, 'R.F. H.T. power supplies for cathode-ray tubes', *A.W.A. Tech. Rev.* 7 (1946), 53.

E.H.T. can be obtained from a low-voltage source without a transformer by employing the Cockcroft-Walton voltage multiplier circuit, illustrated in outline by fig. 7·10. This principle has been used in conjunction with metal rectifiers* for obtaining a 5 kV. E.H.T. supply from a 350–0–350 V. mains transformer winding. The operation is as follows: the rectifier MR_1 charges C_1 to the peak input voltage $\hat{V}_i$ in the usual manner, and the peak voltage applied to MR_2 in series with C_2 is therefore $\hat{V}_i + V_{c1}$ or $2\hat{V}_i$,

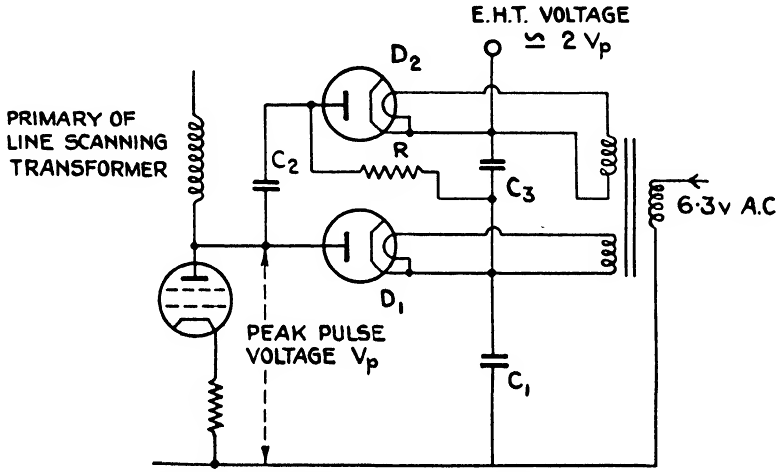


Fig. 7·11. Voltage-doubling circuit using valve rectifiers for obtaining E.H.T. voltage from line fly-back. $C_1 = C_2 = C_3 = 0{\cdot}001\ \mu$F. of working voltage V_p.

so that C_2 is charged to this value. Across the circuit MR_3, C_3, we have a peak voltage $(\hat{V}_i + V_{c2} - V_{c1}) = 2\hat{V}_i$, so that C_3 also charges to this value. Across R_4C_4 we have a peak voltage

$$\hat{V}_i + V_{c1} + V_{c3} - V_{c2},$$

i.e. $2\hat{V}_i$ again. The total output voltage with n rectifiers is therefore $n\hat{V}_i$. If too many sections are used the regulation is not good enough even for the value of low-load current normally involved, and a voltage multiplication of 7 or 8 times is probably a reasonable maximum. A somewhat different circuit† may be used for voltage doubling or tripling, when deriving E.H.T. voltage by means of the line-scan fly-back pulse,

* A. H. B. Walker, 'Television E.H.T. supply', *Wireless World*, May 1948.

† See, for example, W. T. Cocking, 'Television E.H.T. supply', *Wireless World*, June 1947.

and an arrangement in domestic use by the author is illustrated in fig. 7·11. The pulse voltage V_p charges C_1 via D_1 in the normal way, and between pulses C_2 becomes charged from C_1 through R. The voltage applied to the circuit C_3, D_2 during the pulse is therefore $V_p+V_{c2}-V_{c1}$, i.e. V_p, so that C_1 and C_3 are both charged very nearly to the peak voltage of the pulse. Valve rectifiers were used, as they happened to be available at relatively low cost; as in any of the more usual valve rectifier systems it is necessary to have a heater winding capable of withstanding the full E.H.T. voltage, but this presents no great problem and the additional heater winding is a negligible complication since its insulation (and the insulation between windings) need only stand half the voltage. The condensers also need only be rated to stand about half the output voltage. It is to be noted, however, that voltage multiplier circuits of either of the types described lend themselves particularly to the use of metal rectifiers, since if n rectifier units are required each handles $1/n$ of the total voltage, and approximately the same number of rectifying elements is necessary whether we employ a high input voltage and a single rectifier assembly, or a low input voltage and a larger number of smaller units.

The new methods of E.H.T. generation described above probably constitute the most important of recent advances in the television field, judged in terms of cost and engineering convenience. In particular they make it possible to dispense with mains transformers and at least one such receiver has recently appeared on the market.

Chapter 8

SOME NEW KINDS OF RECEIVER

In the early chapters we were concerned mainly with advances in basic ideas. This is only a small part of the story of recent development, which has been largely concerned with the application of well-known ideas to new purposes, and a comprehensive account of all the new kinds of receiver, their engineering problems, and the progress made, would be a rather formidable undertaking. In this chapter a few of the more important items will be selected for brief description and discussion.

8·1. Radar receivers

8·1·1. *General*

It is possible to envisage a very simple kind of receiver for radar, consisting of an amplifier tuned to the transmitter frequency with its r.f. output applied directly to the deflexion plates, or used to vary the brightness of the spot of a cathode-ray tube supplied with a suitable time-base. It is more satisfactory, however, to use a detector followed by video amplification, especially in the case where the brightness is varied; if we were to apply an r.f. (or i.f.) signal directly to the grid of an indicator tube only the positive half-cycles would be effective, whereas to achieve adequate brilliance of the picture it is necessary to apply the brightening impulse for the full duration of the pulse.

In practice superheterodyne receivers are always used, because of the difficulties of constructing or maintaining r.f. amplifiers at the high frequencies employed. The mixer is preceded by one or more r.f. stages when working with frequencies up to 600–1200 Mc./sec., but at higher frequencies the aerial is usually coupled directly into a crystal mixer via a suitable protecting device. The overall band-width for the r.f. and i.f. stages is usually between 1 and 4 times the reciprocal of the pulse length, depending largely on whether detectability of weak signals, or accurate ranging, is the main consideration. The first amplifying valve is usually required to have the lowest possible noise factor, and the overall gain must be sufficient to ensure that the range of reception is limited by noise level and not by lack of gain.

The display is often situated some distance from the rest of the receiver, and use is made of cathode followers which, by virtue of their relatively low output impedance compared with amplifiers, are suitable for feeding into long lines. The single EF 50 used as a cathode follower in figs. 6·14 and 6·16 has an output impedance of 200 ohms and is suitable for feeding about 10 ft. of 70-ohm cable with 1/μsec. pulses. With longer cables or shorter pulses it is usually desirable to present a good match to the cable, and it may be necessary to use two or more valves in parallel. In the first example the voltage applied between grid and cathode of the valve is small compared with the output voltage, but in the matched case the grid-cathode voltage is roughly equal to the output voltage, and the valves used have to have a long grid-base as well as high slope; this involves a relatively high-power dissipation.

8·1·2. *Common aerial working*

It is commonly required to use the same aerial for transmission and reception, and except with very low powers this requires some device to short-circuit the input to the receiver during the pulse and prevent damage. Such a device performs the additional function of preventing the wastage of transmitter power by absorption in the receiver, and is known as a T.R. (transmit-receive) switch; it may employ a spark gap, a gas-discharge device, or a negatively biased diode which is non-conducting for small signals but conducting for large ones. To prevent absorption of some of the received signal by the transmitter, it may be necessary to use a 'transmitter blocker' to short-circuit or disconnect the transmitter during reception.

There are many ways in which such an automatic switching system may be arranged, but the basic principles can be illustrated by reference to fig. 8·1. It will be recalled that a quarter line of impedance Z_0 acts as a transformer such that a low impedance Z_L connected across one end is transformed to a high impedance Z_H at the other according to the relation $Z_L Z_H = Z_0^2$. In reception the open quarter-wave line AB provides a low impedance at B which prevents signal power from being fed into the transmitter. At the same time, since the length BC is a quarter wave-length, the low impedance at *B* appears as a high impedance at *C* and allows signal power to pass unhindered to the receiver. On transmission, conduction takes place at *A* and *D*, so that high impedances are now presented at *B* and *C* and the transmitter power can pass to the aerial without undue loss; at the same time the receiver input is short-circuited.

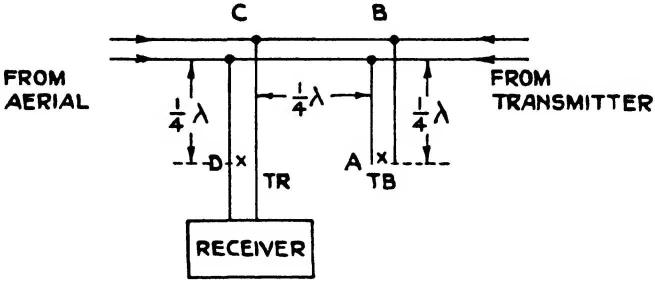


Fig. 8·1. Common aerial system.

The transmitter blocker or T.B. (at *A*) is not always needed, because the transmitter impedance changes considerably when the modulation is applied, so that on 'receive' the transmitter is not matched to the feeder and by adjustment of distance between the transmitter and *C* a high impedance can be presented at *C*, on 'receive', even without the T.B. device. The objection to this is that the impedance at *C* depends on the transmitter tuning and may not be entirely under control. Another function of the T.B. in certain metre wave-length radar equipments arises because it is not convenient to cut off the transmitter valves entirely when receiving, and it is necessary to prevent shot noise due to the residual anode current from being fed into the receiver.

At centimetre wave-lengths the T.R. and T.B.* switches usually consist of resonant cavities incorporated in gas-filled bulbs. In the absence of discharge, signal is transmitted through the cavity with a loss of the order of between 0·5 and 2·0 db. only; the transmitter pulse causes a discharge which damps the cavity and thereby prevents the greater part of the pulse from getting through to the receiver, although there is a small amount of leakage, and also a relatively large initial spike due to the finite time required for

* The term A.T.R. or anti-T.R., though often used in place of 'transmitter blocker', is deprecated.

ionization of the gas. De-ionization of the gas at the end of the pulse is also not instantaneous, and full recovery may take several microseconds, depending on the kind of filling employed.

For a detailed treatment of this subject the reader is referred to the volume of this series dealing with aerials.

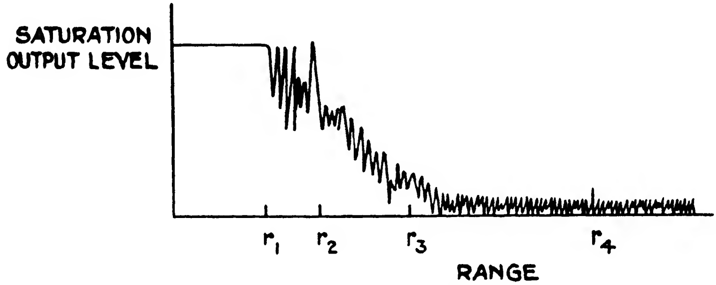


Fig. 8·2. Radar amplitude-time display with clutter.

8·1·3. *The clutter problem—swept gain and I.A.G.C.*

A serious problem in radar receivers is the loss of wanted information owing to saturation by unwanted signals or 'clutter'. Fig. 8·2 shows the appearance of a range-amplitude display with wanted signals at r_2 and r_4 and large numbers of unwanted and overlapping reflexions at ranges up to r_3. Up to the range r_1 there is complete saturation, and any wanted signal is entirely lost. The signal at r_2 is visible, but will be lost if the gain is turned up slightly higher. In the case of P.P.I. or other displays in which the brightness is modulated, the situation is more serious because the whole range from black to white is covered by a variation of amplitude of the order of 2 to 1, and all echoes, however strong, at ranges up to r_3, i.e. so long as the clutter is more than two or three times the mean noise level, are obscured. Wanted echoes up to r_1 can be obtained by turning the gain down, but then the echo at r_4, which may be wanted at the same time, is lost.

Various devices are in use for avoiding saturation by clutter. In certain cases, such as that of reflexions from the waves of a uniformly rough sea, the mean clutter amplitude of the clutter decreases with range according to a definite law, and by making the receiver gain rise from a low initial value according to the inverse of this law it is possible to preserve for all ranges maximum possible visibility of echoes against the clutter or clutter-plus-noise background. This is sometimes known as 'swept gain' and may be achieved by means of a very simple arrangement such as that shown in fig. 8·3, in which part of the modulation pulse of the transmitter is used to charge the condenser C_1, the voltage developed across C_1 being adjusted by means of the resistance R_1 and applied to the grid of an i.f. valve so reducing its gain. At the end of the pulse, the bias decays exponentially at a speed controlled by C_1 and the resistance R_2 connected across it; by suitable choice of R_1, R_2, C_1, an approximation is obtained to the desired law of gain versus range.

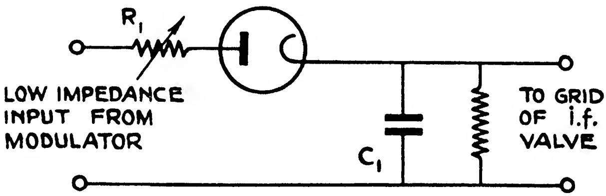


Fig. 8·3. Simple form of swept gain circuit.

Even when the clutter obeys no definite law, swept gain can often help in achieving a useful compromise between the gain requirements for short and long ranges.

Another device which can be useful when circumstances are favourable is quick-acting a.g.c., that is to say, an a.g.c. system with an operation time of not more than a few pulse lengths. This changes the block of clutter plus wanted signal sketched in fig. 8·4(*a*) to the form of fig. 8·4(*b*). Passage of the latter through a short time constant, or differentiating, circuit such as fig. 9·1(*a*) or (*b*) produces the result of fig. 8·4(*c*), the negative impulses being removed by means of a diode connected across the output terminals. A rather better system, relatively free from paralysis troubles, is provided by the logarithmic receivers of § 9·8 used in conjunction with differentiation as suggested by R. V. Alred and A. Reiss.

8·1·4. *Automatic-frequency control.*

Most microwave radar receivers employ automatic-frequency control.

Automatic-frequency control circuits were first described by Round in a patent published in 1921, but the principles did not find practical application until 1936 when they were embodied in certain broadcast receivers. In these original forms of a.f.c. circuit a permissible small frequency error Δf produces a frequency change of opposite sign, and of magnitude $A\Delta f$, thus reducing a tuning error of $(A+1)\Delta f$ to one of Δf, but use can also be made of discontinuous mechanisms whereby a tuning error in excess of a certain amount operates some form of trigger device

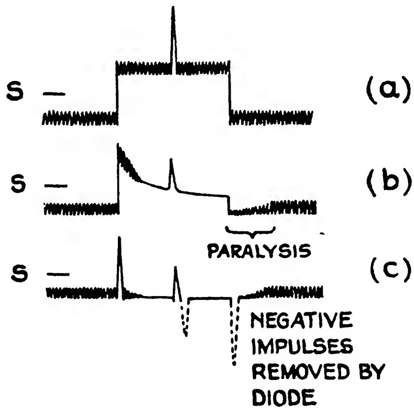


Fig. 8·4. Sketch illustrating quick-action a.g.c. with differentiation, showing an amplitude-time display with a wanted signal superimposed on an idealized rectangular block of clutter. S = saturation level of P.P.I. display.

which pulls the tuning back in the right direction. Such systems may operate in either direction, or the receiver tuning may be given a continuous tendency to drift in one direction, which is repeatedly countered by a 'kick' in the other direction so that the possibility of serious drift is replaced by a continuous hunting of small enough amplitude to be tolerated. A further alternative is some form of motor-driven tuner with the supply to the motor automatically cut off at a point near enough to the correct tuning point.

All a.f.c. systems at present used employ some form of discriminator circuit, which derives from the signal a d.c. voltage with its sign determined by the direction of the tuning error. The output of the discriminator is applied, possibly after amplification, to alter the frequency of the local oscillator by means of some device which may be either electronic or mechanical. Electronic devices include reactor valves, klystrons designed so that the frequency is a function of reflector voltage, and valves of the 'Heil' type with their frequency dependent on anode voltage. Mechanical devices are numerous and include various forms of motor-driven and thermally operated mechanisms which in general offer the advantage of wider frequency coverage at some cost in terms of convenience. All these systems have two main properties requiring separate consideration, namely, the ability to hold the tuning in step once it has been established, and the ability to pull the receiver into correct tune provided the initial error does not exceed a certain amount.

The designer of an a.f.c. system for radar purposes is faced with a variety of considerations which are special to radar; for example, he is helped by the fact that in radar equipments a.f.c. circuits have to tune the receiver to the adjacent transmitter instead of to incoming signals and can therefore be made proof against locking by unwanted signals, provided suitable steps are taken such as the use of a separate a.f.c. mixer which takes its 'signal' direct from the transmitter and is not coupled to the aerial.

On the other hand, there are special difficulties; thus the signal from which the control must be derived exists only for very brief periods at relatively long intervals, but the control voltage must not change appreciably during these intervals, otherwise the frequency of the local oscillator will be different for signals at different ranges and the receiver will be mistuned at the commencement of the following pulse. The alternative possibilities are either to derive the control voltage by integration of pulse energy over a sufficiently long period, or to use pulses to produce a triggering action as described above.

Integration in the discriminator. One method is to carry out the integration in the discriminator circuit. This requires large diode loads with high capacities across them, and entails heavy damping of the i.f. circuits, since enough energy must be derived during the pulse to maintain a steady voltage for a relatively long subsequent period. In order to achieve a satisfactory discriminator response without infringing valve specifications, it is necessary to accept an appreciable voltage loss in the integration process. In consequence it becomes necessary to use d.c. amplification following the discriminator, and this in turn leads to difficulties in maintaining constant operating conditions. To ease this problem, the output of the discriminator should be made as large as possible, and this favours a discriminator employing two valves with their output circuits staggered in tuning rather than the single-valve phase, or 'Foster-Seeley' type of discriminator. A further argument against the phase discriminator for this application is the difficulty of integrating and at the same time maintaining symmetry.

Integration after amplification. Pulses from the discriminator may be passed through a video amplifier and integrated subsequently, or alternatively the integration process may be a partial one producing a saw-tooth output which can be applied to a 'video' amplifier, full integration being carried out subsequently. In these cases phase discrimination is more attractive, since its disadvantages are less troublesome, it may save a valve, and it is easier to deal with if it gets out of adjustment.

Intermediate-frequency components of pulses. The spectrum of radar pulses extends into the i.f. region, and unless precautions are taken this is liable to have a disturbing effect on the system, either as a result of direct pick-up of the modulator pulse in the a.f.c. circuits, or by demodulation of the r.f. pulse at the mixer. Up to a point, these unwanted components can be cancelled out by an appropriate arrangement of circuit. Another difficulty is that the spectrum of the transmitted pulses is liable to contain 'bumps', and a.f.c. systems are apt to lock-on to one of these.

The inherent pull-in properties of a simple a.f.c. system depend on symmetry of the discriminator to a much greater extent than the 'hold-on' properties. The additional amplification required in a pulse a.f.c. system introduces extra tolerances, and symmetry may be further upset in consequence of the effects described above. Local oscillator 'hysteresis' is a further source of trouble, and may result in the local oscillator failing to start. A 'sweep' circuit is therefore essential if a radar a.f.c. system is to be relied on for pulling itself into tune.

Image and harmonic locking. A sweep circuit, in turn, brings in additional problems. Principal difficulties likely to arise are the prevention of locking on image and harmonic responses; fig. 8·5 shows the responses obtained from a typical discriminator as the local oscillator or signal frequency is altered, and the various possibilities of incorrect locking will be readily seen if it is borne in mind that stable operating points can occur whenever the slope of the curve is of correct sign, and sufficiently steep. Both harmonic locking, i.f. signal locking, and (within limits) locking on side peaks of a bad transmitter spectrum can be avoided by ensuring that the wanted response is less than, or not much greater than, that which just saturates the discriminator, so that amplitude discrimination is preserved between the wanted and unwanted responses. As an additional precaution it is desirable to avoid equality of signal and local oscillator voltages in the a.f.c.

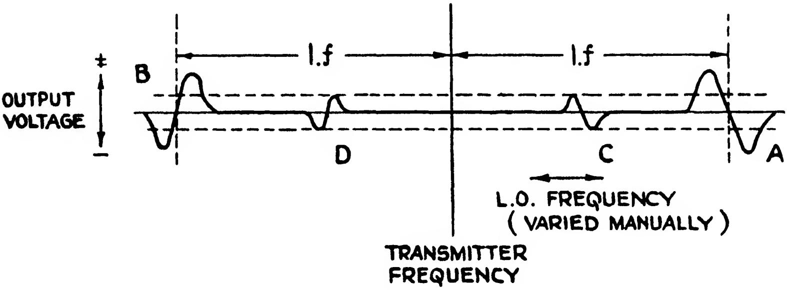


Fig. 8·5. Variation of discriminator output from typical discriminator as the local-oscillator frequency is varied. *A*, wanted response; *B*, image response; *C* and *D*, harmonic responses.

mixer, so that harmonic production is not excessive. Image locking is serious because it occurs just *off* tune, and no simple and completely satisfactory way of avoiding it has so far been evolved except to choose an i.f. greater than half the sweep range.

8·1·5. *Automatic-gain control*

It is sometimes required to hold a radar echo, which may be fading with a period of say 30 c./sec., at constant amplitude. This may be done by means of a locally generated 'gating' pulse which is adjusted to coincide with the echo; echoes occurring simultaneously with the gating or 'strobe' pulse are integrated to produce a steady gain-controlling voltage, and the selected echo is subject to normal a.g.c. action.

8·2. Panoramic receivers

A panoramic receiver is used to give a picture, at a glance, of what is happening over a wide band of frequencies. Such a device can be used amongst other things to find an unused channel for communication, and in a variety of ways for radio navigation.*

Fig. 8·6 illustrates a panoramic receiving system, in block form. The frequency of the local oscillator is swept in more or less linear manner over the frequency range f_1 to f_2 at about 25–50 c./sec., and causes the i.f. amplifier, which is tuned to f_3, to accept signals varying from $f_1 \pm f_3$ to $f_2 \pm f_3$. The spot of a cathode-ray oscilloscope is moved in sympathy with the frequency variation, producing a trace which can be calibrated in terms of frequency. Received signals can be applied either to deflect the spot producing

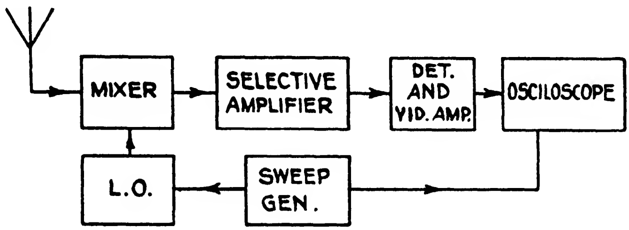


Fig. 8·6. Panoramic receiver.

a picture similar in appearance to a radar amplitude-time display, or alternatively as brightness modulation. The frequency variation can be produced either by mechanical or electronic tuning, but the ranges which can be covered electronically are relatively limited, and any r.f. circuits preceding the mixer must be broad enough to accept the whole of the swept frequency range.

As the frequency is swept through an occupied channel a 'pulse' of signal is produced of duration B/S, where S is the scanning speed in c./sec. and B the band-width of the selective amplifier in c./sec. As in the case of radar, the band-width must be of the order of the reciprocal of the pulse length for maximum detectability of weak signals, giving the relation $B_{(\text{opt.})} \simeq \sqrt{S}$.

Suppose the band to be covered is 10 Mc./sec. and the minimum scanning speed acceptable is 40 c./sec.; then $B_{(\text{opt.})} \simeq \sqrt{(40 \times 10^7)}$ or 20,000 c./sec. As a rule the signals to be received will themselves occupy a much narrower band-width than $B_{(\text{opt.})}$, so that with a normal type of receiver they can be received with much less

* 'Panoramic reception', *Electronics*, 13 (June 1940).

noise; a loss in signal-to-noise is, therefore, a price which usually has to be paid for panoramic reception.

The above discussion applies to signals of a more or less continuous nature, but special problems arise in the reception of pulse signals; in such a case the band-width must be wide enough for reproduction of the pulse, and since pulses are usually only present for a small part of the total time they produce at most a very faint trace compared with the noise background, unless some trick is resorted to such as that of using the signals for brightness modulation as well as for deflexion of the trace.

Various devices such as spiral traces may be used for obtaining a sufficiently open scale to enable the frequency to be read to closer limits, and accurate calibration points may be obtained by injecting signals from a suitable generator of standard frequencies.

Spectrometers for examining the pulses produced by radar transmitters have been constructed on the same principle. In this case the band-width requires to be narrow compared with the width of spectrum to be studied, and since the signal is very strong the amplifier can be simple and there is no background noise problem.

8·3. The crystal-video receiver

The 'crystal-video' receiver is one in which a simple crystal detector is followed by a video amplifier of very high gain. It was evolved in the first instance for a radar beacon system, the requirement being for reception of radar signals—often several at once—anywhere within a band-width of several hundred megacycles, these signals being used to actuate a transmitter sending a pulse, on some other frequency, back to the radar station.

The band-width required was wider than that of any practicable i.f. amplifier, but since the signal strength to be received was relatively large, it was found possible to achieve just adequate sensitivity without using any r.f. or i.f. amplification, and this in turn made it possible to achieve the very wide r.f. band-width required. Such a system is relatively insensitive because the voltage output V_o of a detector operating at low input levels falls off as the square of the input voltage V_i, i.e. directly as the input power. Assuming adequate gain in the amplifier, the limit of sensitivity is reached when V_o is comparable with the equivalent noise input voltage V_n of the amplifier.

The N.T.R. of a crystal tends to unity as the current through it tends to zero; in other words it then behaves to a first approximation just like any normal resistance of the same ohmic value. Under these conditions the resistance R of a silicon crystal is usually of the order of 2000–100,000 ohms, and if it is connected directly in the grid circuit of a video amplifying valve having an equivalent noise resistance R_n, we have

$$V_n^2 = 4kTB(R+R_n),$$

where B is the video energy band-width.

Let us suppose that an r.f. power of b microwatts produces a micro-ampere of rectified crystal current. V_o is then $(R\times 10^{-6})/b$ volts per microwatt, and the signal to noise-voltage ratio for any given input power P_i in the square-law region of the detector characteristic (i.e. up to at least 10 μA. rectified current) is equal to

$$\frac{P_i R\times 10^{-6}}{b.4kTB\sqrt{(R+R_n)}}.$$

Eliminating constants, we are left with $R/b\sqrt{(R+R_n)}$ or approximately $\sqrt{R}/b$ as the factor of merit for the crystal. b depends on several different properties of the crystal; the conductance $\omega^2C_c^2R_b$ (p. 61) again comes into the picture, and this time it absorbs nearly all the available r.f. power, because the linear resistance R is relatively large, being of the order of perhaps 100 times $1/\omega^2C_c^2R_b$. The spreading resistance R_b and contact capacitance C_c therefore determine the amount of r.f. power required to develop a given voltage across the crystal, and if as usual there are no r.f. tuning adjustments, any departure from the value of $\omega^2C_c^2R_b$ assumed in designing the system will produce a mismatch loss, and variation of C_c may produce further loss due to mistuning. For a given applied voltage the output voltage depends on the curvature dI/dV of the rectifier characteristic (fig. 8·7) for the chosen

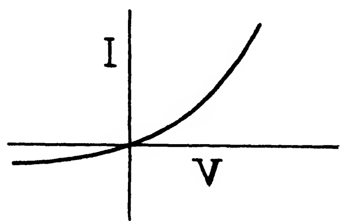


Fig. 8·7. Crystal rectifier characteristic.

operating point which is usually at the origin. It might be expected that a point on the characteristic more suitable for rectification could often be found with the aid of a variable bias voltage; this is true, but unfortunately as soon as current flows through the crystal a lot more noise is generated at video frequency and the net result is usually worse performance, although a small bias voltage, say +0·1 to +0·3 V., effects an improvement with some crystals.

The curvature at the origin is a very variable feature of crystals, but in the mixer case this is not important, since it is, broadly speaking, the *average* forward slope which matters and this does not vary from crystal to crystal to nearly the same extent. In the present state of development, a crystal video receiver performance comparable with the best obtainable requires rather rigorous crystal selection tests.

With good crystals the value of R is always fairly high, and the factor of merit approximates to $\sqrt{R}/b$ provided valves of low E.N.R. are used. If it is required to receive very short pulses, however, care must be taken to avoid too high a value of R, since a minimum practical value for the output capacity of the crystal holder plus the input capacity of the valve is about 15 pF., and with a 30,000-ohm crystal this gives a time constant of 0·45 μsec., seriously lengthening pulses of less than a microsecond or so duration, and also since the capacitance is in shunt with R it reduces the effective factor of merit by decreasing the output impedance for all but the lower frequencies in the pulse spectrum.

The crystal-video receiver is rather difficult to compare for sensitivity with a superheterodyne. The noise-factor concept is not applicable to it because virtually all the noise originates *after* detection. Although the r.f. band-width may be several hundred or even thousands of Mc./sec. and the video band-width only a megacycle or so, the detection efficiency is so low that r.f. noise is not significant; also since the detector is square law the apparent noise factor is inversely proportional to signal strength. There is a scarcity of reliable data on the absolute sensitivity of crystal-video receivers, but figures of merit for good crystals are of the order of 100 or more, and performance is in the region of 30–50 db. down on the superheterodyne at minimum signal levels; for example, if we take unity signal-noise ratio and a figure of merit of

100 we find that the input power $P_t = 4kTB \times 10^4$ and the corresponding value of P_t for a superheterodyne receiver of noise factor N is $NkTB$. If $N = 8$ the advantage in favour of the latter is $0{\cdot}5 \times 10^4$ or 37 db.

The video amplifier, illustrated in fig. 8·8, is of very simple design. The main problem is microphony, and although this can

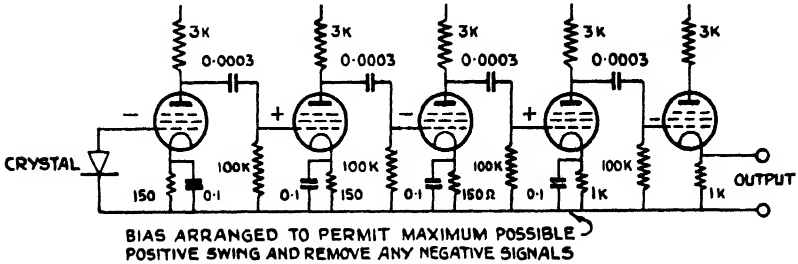


Fig. 8·8. Typical crystal-video receiver.

be kept under control in the laboratory and perhaps for some external applications by suitable choice of valves, and by sound insulation, it is usual to avoid it whenever possible by not allowing

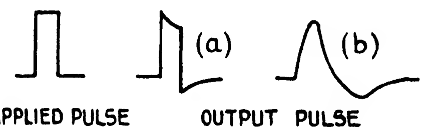


Fig. 8·9. Illustration of "overshoot" in a video amplifier.

low frequencies to pass through the amplifier. This is done in fig. 8·8 by using an intervalve coupling of fairly short time constant so that below about 30 kc./sec. the amplifier response is falling off steeply. A further problem arises out of this, however, because the coupling condensers charge up to some extent during the pulse and gradually lose their charge at the end of it, producing 'overshoot' as shown in fig. 8·9 (*a*) or more usually (due to non-amplification of the higher frequencies) fig. 8·9 (*b*). An exaggerated form of this effect is obtained on strong signals when a grid is driven positive so that the capacity is charged up rapidly, via the low grid-cathode circuit impedance. The overshoot from a strong signal may well cut off one or more of the amplifying valves for some time—perhaps many times the pulse length—to the extent that weaker signals are lost.

8·4. Receiver for measurement of thermal radiation

The question of the equivalent noise temperature of aerials has been briefly discussed in an earlier chapter, and it has been mentioned that the noise temperature of a microwave aerial with its beam directed upwards may be only just above absolute zero. Radiation from an aerial travels outwards until it is ultimately absorbed somewhere, either by terrestrial matter, in the Galaxy, or at even greater distances; conversely, the absorbing matter sends out thermal radiation in accordance with its temperature, the average value of this being the 'equivalent noise temperature' of the aerial*—in the absence, that is, of any further noise such as may be produced, for example, by thunderstorms, by human activities, or by the release of atomic energy in the sun and stars.

Given a receiver sufficiently sensitive to enable us to observe small changes of aerial 'temperature', a variety of interesting measurements becomes possible. For example, suppose the microwave beam of the previous paragraph is 'intercepted' by a rainstorm 3 miles thick at a temperature of 300° absolute; if the noise temperature of the aerial is 150° absolute with the rainstorm present, and 0° without it, we can deduce that the rainstorm would intercept half the energy travelling in either direction between the aerial and any point beyond the rainstorm. The attenuation due to the rain at the wave-length in question is therefore 1 db./mile.

R. H. Dicke has described† a receiving system for the measurement of atmospheric attenuation and of radiation from the sun and moon. His method is a particularly ingenious application of the principles of integration mentioned in § 2·7·1, where it has been explained that the sensitivity of a receiving system can in principle be increased indefinitely provided the time available for observing the signal is unlimited. For purposes such as those mentioned above the 'signal' consists of small changes of noise level, and the time available varies from minutes to perhaps hours or even days. It has not yet been found practicable or necessary to

* The proof of this is very simple. Consider a resistance connected to an aerial, such that noise power from the resistance is radiated by the aerial and absorbed in matter at the same temperature as the resistance; the second law of thermodynamics requires that there shall be neither gain nor loss of energy and therefore the received and transmitted noise powers must be equal.

† *Rev. Sci. Instrum.* 17 (July 1946), 268.

employ such long periods as this, but if we take a receiver of band-width 10 Mc./sec., and increase the observation time from $\frac{1}{10}$ μsec. up to 1 sec. the improvement in the ratio of minimum detectable signal power to noise power will be $\sqrt{n}$ (where n is the ratio of integration time to the reciprocal of the band-width) or about 3000 times. If the source temperature is 300° and the noise factor 10, i.e. if the total noise is equivalent to that of a noiseless receiver coupled to an aerial at 3000° absolute, the fluctuation noise in the output is now equivalent to a temperature fluctuation of only about 1° C. Unfortunately, in practice we only gain from an increase in time of integration so long as the residual noise fluctuations are large compared with the variations of output due to random changes of receiver gain. Such changes originate in valves and other components of the receiver, even if all the supply voltages are accurately stabilized and the temperature maintained constant; as the integration time is increased, these changes can build up to larger amplitudes and at the same time the amount of fluctuation which can be tolerated is decreased. In the above example we have supposed the detection of a change in r.f. signal power of only 1 part in 3000, which amounts to only 0·03% of the amplitude following detection by a linear rectifier. To achieve this degree of stability, even over periods as short as 1 sec., is rather difficult. Dicke's method consists of measuring the aerial temperature relative to room temperature at a frequency of, say, 30 times a second, and then taking the average of a large number of these observations. The receiver cannot drift nearly so far in $\frac{1}{30}$ sec. as it can in, say, 1 sec. and the background due to the receiver fluctuations is reduced accordingly.

The way in which this is done is illustrated by fig. 8·10. The aerial is coupled to the receiver via a length of slotted wave guide. The disk consists partly of absorbing material and rotates at 1800 r.p.m., being so designed that it would produce very nearly square-wave 30 c./sec. modulation of an incoming signal. If the aerial noise temperature is the same as that of the disk (and provided care is taken to ensure that any variable r.f. mismatch introduced by the disk is either small or does not affect the receiver noise level), the mean output of the receiver does not vary as the disk rotates. Any difference appears as a 30-cycle modulation of the output from the second detector, and after passing through a selective 30 c./sec. amplifier it is mixed with the output of a 30 c./sec. generator mounted on the same shaft as the disk. The amplifier output due to the wanted modulation is either in phase or 180° out of phase with the generator output, depending on whether the aerial noise temperature is higher or lower than room temperature, and produces either an increase or decrease of d.c. voltage at the output of the mixing valve. The random variations are equally likely to affect the output in either

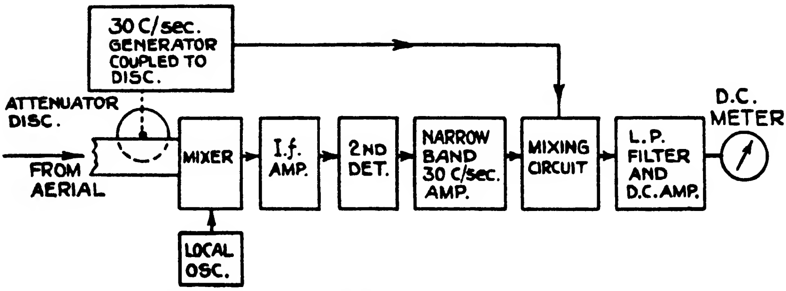


Fig. 8·10. Receiver for measurement of thermal radiation in the micro-wave region, using post-detector integration.

direction, so that the resultant tends to zero as the time constant of the meter indicating the output of the mixing valve is increased. The combination of low-pass filter and d.c. meter is designed to have the properties of a long time constant, critically damped, galvanometer.

With a meter-plus-filter time constant of 2·5 sec. and an i.f. band-width of 16 Mc./sec., Dicke has been able to detect a temperature change of 0·5° C.

It will be clear from § 2·7·1 and the example given above that every time we double the r.f. band-width the sensitivity is increased by 1½ db. provided everything else including the noise factor stays constant. Unfortunately, if the band-width is increased above about 10 or 15 Mc./sec. it will be found in practice that the deterioration of noise factor more than offsets the increased effectiveness of integration.

8·5. The Synchrodyne

The superheterodyne receiver is so firmly established for most purposes that the recent article by Dr D. G. Tucker on the Synchrodyne (*Electronic Engng*, March 1947) comes as something of a shock, and although tranquillity may be restored somewhat by regarding the Synchrodyne as a superheterodyne of zero intermediate frequency (with questionable use of terminology!), or as the old-fashioned homodyne with its defects eliminated, the idea is both novel and extremely interesting.

Essentially the Synchrodyne consists merely of a local oscillator locked to the incoming carrier, plus a suitable mixer to which the signal and the local oscillation are applied and which provides an audio output. To compete in sensitivity with the superheterodyne a pre-amplifier is required, which can in principle be untuned, although some degree of tuning may be necessary to prevent cross-modulation between strong incoming signals. To eliminate heterodyne whistles and reduce 'side-band splash' a filter may be added to the output.

The principle of operation is as follows. The mixer is of multiplicative type, that is to say, the local oscillator is used to modulate the signals so that the output voltage at any instant is determined by the product of the local oscillator voltage and the signal voltage. The output of the mixer includes beats between the local oscillator frequency on the one hand and, on the other, all the signal carrier and side-band frequencies which may happen to be present. Since the modulating oscillator is locked to the carrier of the wanted station these beats include the desired audio-frequencies. The only other beats in the audio range will be heterodyne whistles and those forms of side-band splash which involve the wanted carrier, and if troublesome these can for the most part be removed by a low-pass filter. With the important proviso that the modulating device and all preceding circuits must be linear with respect to the applied signals, there will be no modulation of one signal by another and unwanted signals will not be demodulated.

There is no great difficulty in designing a conventional superheterodyne to achieve comparable selectivity, and a broadcast receiver with 8 kc./sec. band-width, three pairs of coupled circuits and possibly a whistle filter, will rarely suffer from inter-station interference other than 'splash'. On the other hand, a steeper-sided response curve would enable the fidelity to be improved by increasing the band-width say to 12 kc./sec. for the same degree of freedom from interference. At i.f. this would be too elaborate and costly a procedure for general use, but the l.f.

filter required for obtaining similar performance with the Synchrodyne circuit is relatively simple and cheap. The effective carrier reinforcement provided in the Synchrodyne should be useful in combating selective fading* provided the carrier does not fall below the locking level.

Fig. 8·11 illustrates a typical Synchrodyne receiver in block form. The signal voltage required for locking the oscillator is of the order of 0·07 V., and this determines the amount of r.f.

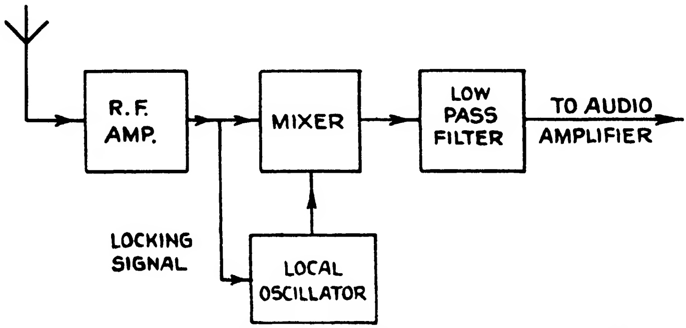


Fig. 8·11. Block diagram of typical Synchrodyne receiver.

amplification required. The mixer requirements are satisfied in principle by the familiar triode-hexode or by a pentode with the local oscillations fed on to the suppressor grid, but the importance of avoiding cross-modulation makes it preferable to employ a mixer such as the Cowan modulator illustrated in fig. 8·12. During one-half of the oscillator cycle the rectifiers become low resistances and the signal path is short-circuited, thereby modulating the signal. From the point of view of the signal channel we have two pairs of rectifiers connected back-to-back; if a signal voltage is large enough to decrease the impedance of one rectifier the effect will be offset by an increase in the impedance of the other rectifier of the pair, thus preventing cross-modulation except in the case of very strong signals. If signal voltage could be developed across the oscillator terminals of the network, cross-modulation would result, but provided the characteristics of the rectifiers are well matched the arrangement can be regarded as a balanced bridge so that the application of a voltage to either pair of terminals

* For a description of earlier proposals for this purpose see § 9·6.

does not result in a voltage at the other. The Cowan modulator requires a low impedance input, e.g. from a cathode-follower.

Two practical difficulties require to be mentioned. If continuous tuning is provided the oscillator will become unlocked between stations and beat with the neighbouring carriers to produce an unpleasant whistle. Against this it is to be noted that distorted reception due to mistuning, one of the serious problems

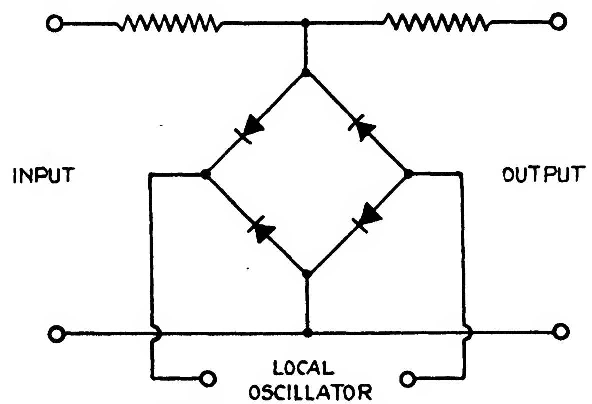


Fig. 8·12. Typical synchrodyne mixer based on the Cowan modulator.

in conventional receivers, is impossible, and it ought to be possible to eliminate the whistles by some variant of interstation noise-suppression technique. The value of continuous tuning is apt to be overrated, and the problem of the whistles does not arise if some form of switched station selection is employed.

Another problem, which does not seem as yet to have received much consideration, is the reception of a weak signal adjacent to a very strong one. As an example let us consider the case of a 0·2 mV./m. signal 20 kc./sec. off tune from a 200 mV./m. signal. Supposing we require 0·07 V. for locking the oscillator and the mixer will accept up to 3·5 V. without overloading or cross-modulation, the selectivity required is 1000/50 or 20 to 1; at 1 Mc./sec. this could be realized with two r.f. tuned circuits, although some degree of compromise between selectivity and quality would be necessary. The selectivity required in the case of superheterodyne reception would be about 500 times greater and is typical of what is achieved in conventional broadcast receivers with a similar degree of compromise.

Chapter 9

SOME NEW CIRCUIT TRICKS

Recent years have seen the evolution of a very large number of novel circuits, and to cover this ground fully would require many volumes. The following examples of new circuit tricks are intended to be as representative as possible, but such a selection must inevitably be a personal one biased to some extent by the authors' own experience and interests.

9·1. Differentiating circuits

Reference has already been made to the use of differentiation, in conjunction with quick-acting-automatic gain control, for reducing clutter in radio receivers. There are many other uses of this idea; for example, it may be required to generate two pulses separated by an adjustable time interval, and this can be done with

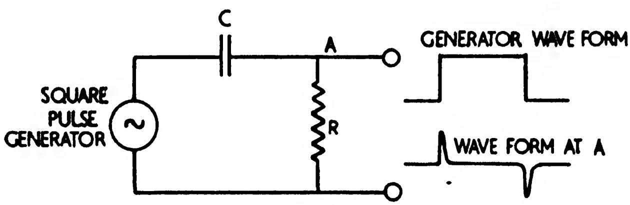


Fig. 9·1 (*a*). Simplest form of differentiating circuit.

a square pulse of variable width obtained from a conventional multi-vibrator. If this is applied to a short time constant, or differentiating, circuit of the type shown in fig. 9·1 (*a*), a pulse appears at *A* for only so long as is required for the condenser *C* to charge up to the amplitude of the pulse. Similarly, at the end of the square pulse the sudden disappearance of the applied voltage means that the potential of *A* drops by an equal amount, thus giving rise to a pulse of opposite sign which lasts until *C* has had time to discharge. If required, the pulses can be converted to a rectangular shape by amplifying them until they are large compared with the grid base of a valve to which they are applied. Limiting at the two ends of the valve characteristic chops off the tops of the pulses leaving them more or less rectangular. The pulses may be separated and used for different purposes; for example, by applying them to the grids of two valves, one grid being biased negatively so that it can only be driven positive, and the other with normal bias and a diode arranged to prevent it being driven positive.

A particularly interesting form of differentiating circuit* is illustrated in fig. 9·1 (*b*). A square pulse applied to the input of the delay line is cancelled by the reflexion from the far end of the

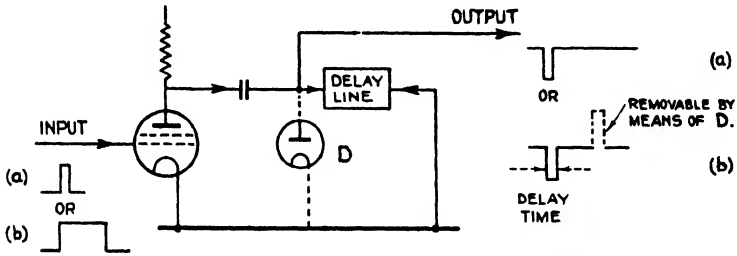


Fig. 9·1 (*b*). Delay line differentiation. The reflected pulse, not shown in Case (*a*), is removable by means of *D*.

line after the appropriate delay time. If the total delay time is equal to or greater than the width T_p of a wanted pulse signal, there will be no distortion or loss of amplitude. The last portion (width T_p) of the reflected signal is uncancelled but inverted, and can therefore be removed by means of a diode as illustrated. Any signal of constant amplitude but longer duration is eliminated except for its first portion which appears as a pulse of length equal to the delay time.

9·2. Amplitude limitation in video amplifiers

Valves which saturate at some relatively low level of applied signal and give a constant output for all stronger signals have been used for many years in the i.f. amplifiers of f.m. receivers.

In pulse receivers, use is commonly made of limiting circuits in video amplifiers, and an example of this is illustrated in fig. 9·2. The object in this particular case is to prevent defocusing of the

* E. Parker and J. E. Keyston, 'Improvements in and relating to gating circuits for the reception of pulse signals': British Patent Application No. 23362/46, quoted by R. V. Alred, *J. Instn Elect. Engrs*, 93, part IIIA, no. 10 (1946), p. 1597.

c.r.t. spot, and possible damage to the tube, which could result from the application of excessive voltage to the grid. The signal voltage from the video amplifier V_1 is applied to the grid of V_2, which is so arranged that it can only be driven negative. The voltage levels, a, b, correspond to the grid base of V_2, so that the signal amplitude is restricted as shown. About half the limiting level corresponds to black, and the limiting level to white.

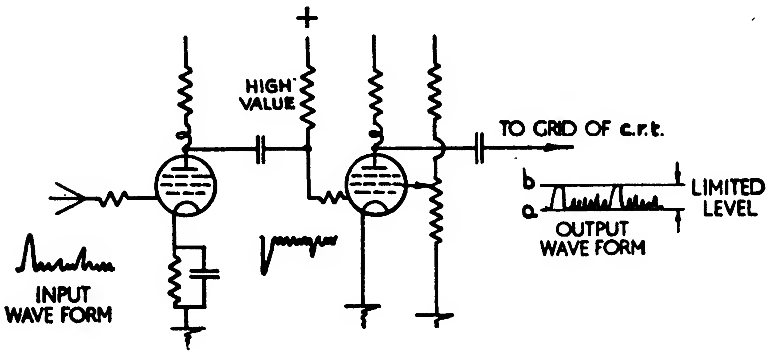


Fig. 9·2. Video limiter.

9·3. Cathode followers of very low output impedance

A main feature of the usual type of cathode follower is the low output impedance which is approximately equal to $1/g_m$, where g_m is the mutual conductance of the valve. This result may be derived very simply by supposing a voltage change Δv at the cathode; since this exists between grid and cathode it produces a cathode current change Δi equal to $\Delta v g_m$, and the impedance looking into the output terminals is given by $\Delta v/\Delta i$ or $1/g_m$.

If an even lower impedance is required, the mutual conductance may be effectively increased by means of auxiliary amplifying and phase inverting stages as shown in fig. 9·3. If the voltage gain of the additional stages is A, the output impedance is given by $1/Ag_m$, using precisely the same line of reasoning as in the simple case. It is possible to dispense with the phase-inverting stage by providing feed-back from the cathode of the last valve to the grid of the first, the cathode of the first stage being earthed. Such circuits may be used to provide correct matching to low impedance lines. If desired, very low output impedances, say 1–2 ohms, are readily obtainable.

9·4. Automatic gain equalization of two receiving channels

The technique described in the following paragraphs was evolved in the first instance for radar purposes, but the idea can be applied, in principle, to equalization of the gain of any two receiving channels provided that brief interruption of reception, at regular intervals, can be tolerated.

A locally generated r.f. signal is fed through equal paths into the two channels, during the intervals. (In radar these exist between the end of one trace and the start of the next.) This

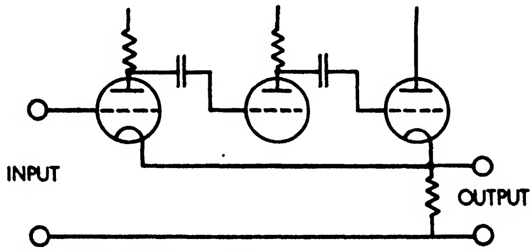


Fig. 9·3. Cathode follower of very low output impedance.

signal is used to operate a.g.c. systems on each of the receiving channels, the two a.g.c. rectifiers having their threshold level determined by a common 'delay' voltage (fig. 9·4).

Assuming that the signal generator or 'equalizer' injects equal signals into the two aerials or feeders, of sufficient amplitude to bring the a.g.c. into operation, and that normal signals are of too short a duration to influence the a.g.c. (or are in some way prevented from doing so) the gain of each chain will be given respectively by

$$G_1 \simeq \frac{V_D + V_1}{V_S} \quad \text{and} \quad G_2 \simeq \frac{V_D + V_2}{V_S},$$

where V_D is the delay voltage, V_1 and V_2 the amount by which the signal outputs must exceed V_D in order to give requisite amount of control, and V_S the equalizing signal voltage at the input to the receivers. Provided $(V_1 - V_2)$ is small compared with $(V_D + V_1)$, then $G_1 \simeq G_2$. No difficulty exists in making V_D large enough to justify this assumption, but the assumption of short duration of signals may be upset in the case of radar if a large amount of ground clutter is present on one receiver and not on the other; this is readily overcome by suitable gating.

An alternative arrangement requires one receiver to have initially a higher gain than the other. Automatic-gain control derived from the difference in the outputs of the two receivers may then be used to bring the gain of the first down very nearly to that of the other.

Both of these systems were tried out at i.f., and satisfactory control obtained in each case over a sufficient range (± 2 or 3 to 1) to correct any likely accidental gain variations; unfortunately, this

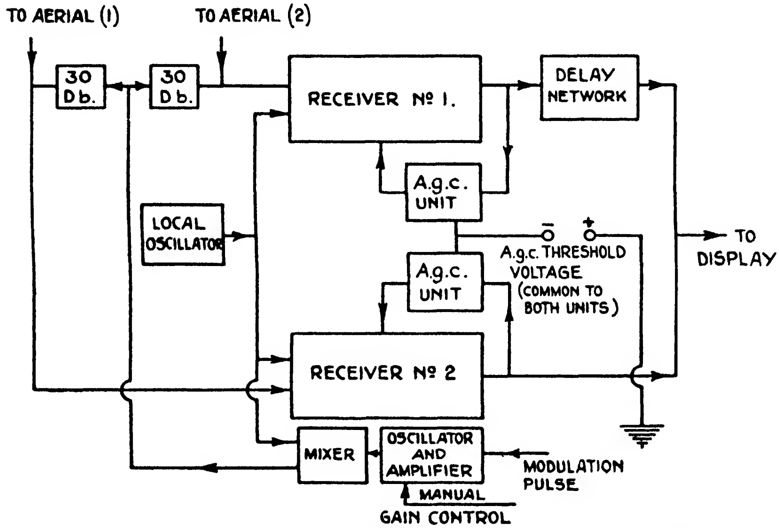


Fig. 9·4. Block diagram of automatic gain equalization system for radar. The equalizing signal is generated at i.f. and mixed with the local oscillator to obtain r.f.

was not the only problem and a serious obstacle was presented by the requirement for remote manual, or possible automatic control of gain over a range of at least 40 db. without disturbance of the equalization. Gain variation with the first method required change of either V_D or V_S; V_D was easily varied but could only be used for control over the very limited range permitted by the necessity of keeping $(V_1 - V_2)$ small compared with V_D. Change of V_S required some form of remotely controllable r.f. attenuator and was not at first sight very practicable. With the alternative system the gain of channel B can be reduced in the normal manner and that of A should follow it down so long as a sufficiently small gain difference can be made to provide the needed control voltage, but in addition to various practical difficulties which might perhaps have been overcome, a fundamental objection was found to this scheme. As the gain is reduced the equalizer signal stays constant; it may therefore be very different in level from the wanted signal at the second detector, and equalization at the one level is unlikely to correspond with equalization at the other. Hence it was an important design consideration that the wanted signal and the equalizer should be of comparable magnitude, and variation of equalizer signal strength was therefore an essential for manual gain control in excess of a few decibels.

Consideration was therefore given to the possibility of devising r.f. attenuating networks using elements whose characteristics could be altered by the application of d.c. Such elements include thermistors, bolometers, and crystal valves, and the last two of these were tried experimentally; the best result obtained was 15 db. change of r.f. level from application of bias to a pair of crystals.

A much more satisfactory solution to the problem was ultimately found by generating an i.f. signal, passing it through one or two stages of gain-controlled amplification, and mixing it with a local oscillator to provide the r.f. signal. Fig. 9·4 is a block diagram of this scheme in practical form, omitting various complications not essential to the technique.

9·5. Reduction of cross-modulation in communication receivers

The fields of broadcast and communication receiver technique, for reasons of age, have comparatively little to show in the way of new devices although engineering progress continues apace. One outstanding problem of great importance which has recently received a good deal of attention* is the reduction of the cross-modulation caused by powerful signals at the grid of the first valve. If there is a powerful transmitter in the vicinity of the receiver it may produce enough voltage swing at the grid of the first valve to modulate other signals over a substantial range of receiver tuning. The use of an adequate number of tuned circuits in front of the first valve in order to prevent this is liable to be cumbersome, and may cause serious loss of signal strength. One

* Discussion before the Radio Section, 6 Feb. 1946, *J. Instn Elect. Engrs*, 93, part 1 (Sept. 1946).

line of attack is the counter-modulation scheme due to E. Hudec.* The audio-frequency component of the valve anode current due to the interfering signal is allowed to produce a voltage across a cathode resistance which is by-passed for r.f. but not for audio-frequency. This produces modulation of the wanted carrier in antiphase with the original cross-modulation, and by adjustment of the cathode resistance a substantial reduction of the cross-modulation is obtainable.

Another method consists of critical adjustment of bias on the first valve. The idea is to move the operating point to the position on the valve characteristic where the coefficients of the odd-order derivations above the first, averaged over the whole of the potential swing, are as small as possible. With this method the cross-modulation from a 1 V. unwanted signal 1% off time, with a single pre-selector circuit, may be reduced to negligible proportions although at some cost in terms of S/N ratio. It may be used in conjunction with counter-modulation.

9·6. Reduction of selective fading

Selective fading is perhaps the main outstanding obstacle to high-quality transmission of speech and music over long distances. It is usually present to a greater or less extent in short-wave broadcast reception, and many wartime listeners to B.B.C. stations are aware that it can also be a normal feature of medium-wave transmissions, the cause being interaction of signals arriving over paths of different lengths. Consider, for example, two paths of equal efficiency and differing in length by any odd number of half wave-lengths; complete cancellation of the carrier will result, but the side-bands remain. Even partial cancellation can result in violent overmodulation and severe distortion and fig. 9·5 illustrates some typical cases.

So-called 'ordinary' fading (unless very severe) can be countered by a.g.c., and receivers often incorporate an efficient a.g.c. system for this purpose. If the carrier and side-bands are obliging enough to fade up and down together this is a useful device; if, on the other hand, the carrier fades leaving the side-bands, the a.g.c. brings the carrier back approximately to its original level and there is therefore an increase of volume, accompanied by distortion due to

* *Electrische Nachrichten-Technik*, 20 (1943), 123–35.

phase shift and excessive amplitude of the side-bands. In the absence of a.g.c., although the distortion still occurs, it is accompanied by a useful reduction in volume or, at worst, no increase. The simplest measure against selective fading seems to be the use of 'poor' a.g.c., and it is usually possible to maintain audibility during ordinary fading without noticeable accentuation of selective fading. However, more fundamental attacks on the problem have been made recently,* based on the principle of boosting the carrier without the side-bands so as to keep the modulation percentage low.

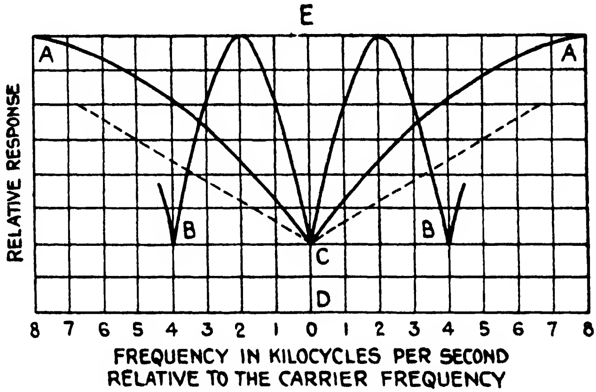


Fig. 9·5. Showing relative response to carrier and side-band frequencies for large and small assumed differences in path length. Difference of 75 km. gives *BCB*; 18·75 km. gives *ACA*.

Experiments were carried out by the author in 1939–40 with the rather crude system shown in fig. 9·6 which had the response curve of fig. 9·7. Artificial but realistic selective fading was produced by using the receiver as the balance indicator of a selective bridge network to which an undistorted medium wave signal was applied, the 'fading' characteristic so obtained being illustrated by the dotted curve in fig. 9·5. The circuit of fig. 9·6 was a complete cure for the artificial fade and rather less effective against medium-wave selective fading. Tests on s.w. were inconclusive. Reliable observation of small improvements on s.w. was impossible owing to the brief duration of the fades, but the main difficulty was attributed to the extremely peaky frequency

* L. A. Moxon, 'Minimizing selective fading', *Wireless World*, Aug. 1941. Murray G. Crosby, 'Exalted carrier reception', *Proc. Inst. Radio Engrs, N.Y.*, 33 (Sept. 1945).

response (corresponding, for example, to curve *B* in fig. 9·5) which appeared to be comparable in unpleasantness with the harmonic distortion which it had been hoped to cure. The absence of

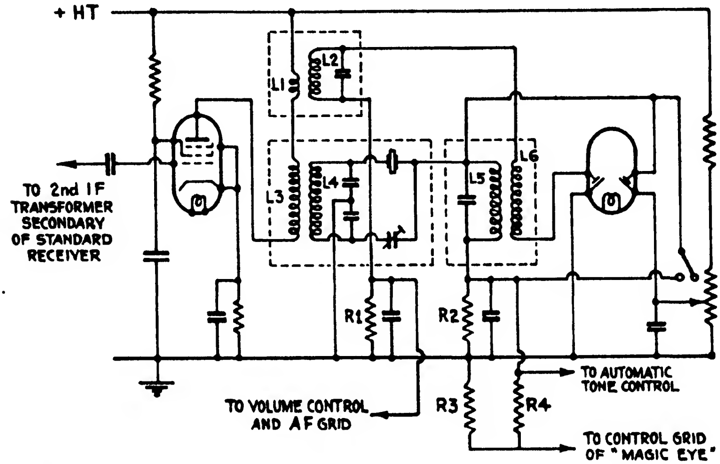


Fig. 9·6. Carrier-boosting circuit. Values of components: L1, small 8-turn coupling coil; L2, 90μH, shunted by 1350$\mu\mu$F; L3, L4, L5, L6, 860μH; R1, 0·2 megohm; R2, 0·5 megohm; R3, R4, 2 megohms; R5, 50,000 ohms; R6, 10,000 ohms.

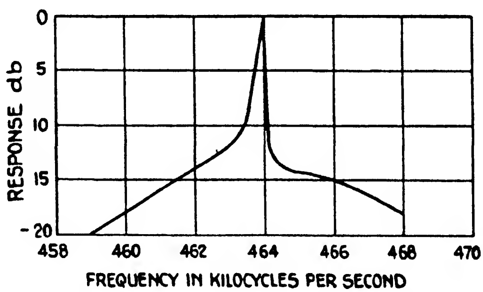


Fig. 9·7. Carrier accentuation: measured response obtained with circuit of fig. 9·6.

an a.f.c. system may also have led to some confusion of the issue, in view of the sharpness of the peak in fig. 9·7. A very noticeable improvement was obtained in many instances, however, when the carrier accentuation was combined with an automatic tone-control action designed to reduce high note response during fading of the carrier. Another difficulty was the accentuation of low-frequency response in the absence of selective fading, and it was necessary to use a fairly drastic l.f. response cut throughout the experiments. The investigation had to be abandoned at an early stage owing to the priority of the war effort. M. G. Crosby has developed the idea much further, and the essential differences between his system and the one described above appear to be in his use of a.f.c. and a more efficient method of carrier emphasis, as illustrated in fig. 9·8.

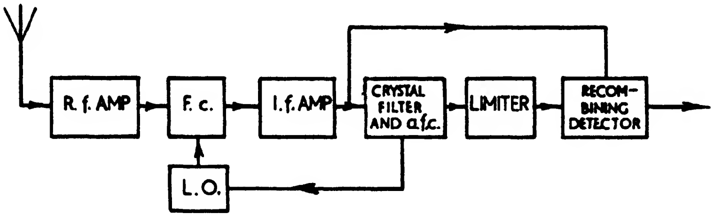


Fig. 9·8. Exalted carrier system due to M. G. Crosby.

The improvement obtained was very considerable, especially in the case of diversity reception; the frequency response difficulty does not appear to have been troublesome, possibly because the improved method of carrier emphasis does not accentuate the lower frequency side-bands. With either arrangement one would expect a selective fade to be accompanied by a loss of the low-frequency side-bands.

9·7. Some new frequency-modulation detectors

In a conventional F.-M. receiver a limiting stage is usually included in the i.f. amplifier and followed by a double-diode frequency discriminator. A most ingenious alternative has recently been described,* in which a single heptode valve combines the functions of limiter, discriminator and amplifier.

Fig. 9·9 illustrates the type of circuit employed. The heptode is made to oscillate at approximately the right frequency, causing the cathode current to take the form of short pulses. The proportion of any one of these flowing to the anode depends on the control grid potential at the time of the pulse, which depends in turn on the phase of the signal relative to the oscillator. The coil in the anode circuit is coupled to the oscillator circuit so that there

* William E. Bradley, 'Single-stage F.-M. detector', *Electronics* (Oct. 1946).

is a transfer of reactance which increases as the anode current increases, and the coupling is chosen so that the oscillator tends to be locked into a fixed phase relation with the incoming signal. As the frequency of the signal varies, there is a corresponding variation of the amount of reactance transfer required, and the

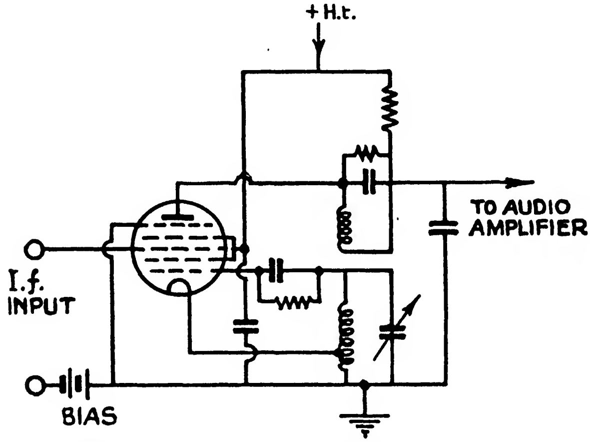


Fig. 9·9. F.-M. detector and limiter using locked oscillator principle.

anode current varies accordingly. The variation of anode current through the resistance R_a provides the audio signal required, and it is stated that an output can be obtained (at full deviation) of 20 V. peak-to-peak, the response to amplitude modulation being 50 db. down on the corresponding degree of frequency modulation. The minimum input signal required to ensure locking of the oscillator up to full deviation is of the order of 0·5 V. r.m.s.

Another important development* is the 'Ratio Detector'. This can be regarded as a modification of the conventional phase (or 'Foster-Seeley') discriminator, and it is similar in action to the extent that the output is proportional to the difference in magnitude of two vectors representing the primary plus half the secondary voltage of a pair of coupled circuits, and the primary minus half the secondary voltage respectively. The special feature of the ratio detector is the provision for keeping the scalar sum of the two vectors constant in the presence of any amplitude modulation. In principle the ratio of the voltages is a function of

* S. W. Seeley and J. Avins, *R.C.A. Rev.*, June 1947.

frequency only, so that if the sum is held constant, the individual values become dependent only on frequency and not on amplitude; in other words, amplitude modulation is removed and a limiter becomes unnecessary. The i.f. gain required is relatively small, 10 mV. or so applied to the grid of the valve which drives the ratio

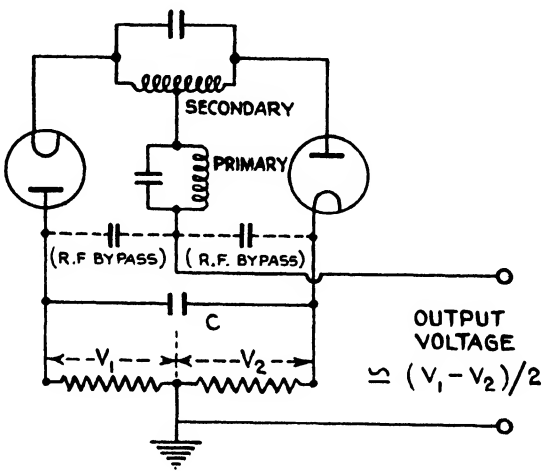


Fig. 9·10. Simplified circuit of ratio detector.

discriminator being a sufficient signal, in contrast to the several volts which may be required to drive the limiter preceding the conventional discriminator.

Fig. 9·10 illustrates the basic circuit. The diodes produce voltages V_1 and V_2, and the sum of V_1 and V_2 is held constant by the condenser C which is made large enough to prevent any appreciable change in voltage over a modulation cycle. The voltage at the output terminals is given by $\frac{1}{2}(V_1 - V_2)$.

The operation of the ratio detector is very much more complex than might appear from this simple description. If the input signal increases the condenser C starts to charge up to the appropriate higher voltage; this lowers the effective load impedance, produces heavier damping of the primary and secondary, and has the desired effect of preventing the output from rising. Conversely, if the input signal decreases, damping is reduced. The action is dependent on the ratio of loaded to unloaded Q of the tuned circuits, and is further complicated by the non-linearity of the diodes and variations of their input reactance with signal level. The damping varies over the F.-M. cycle as well as with amplitude modulation, and the discriminator characteristic depends on the complex impedances of the circuits which in turn depend on the damping. For a detailed analysis the reader is referred to the paper by Seeley & Avins on which is based the following summary of some of the more important practical aspects:

(1) Due to the low input impedance of the diodes, and the relatively high impedance load desirable for the preceding amplifier valve, suitable impedance transformation should be provided.

(2) Optimum rejection of amplitude modulation requires that only a fraction of the total rectified voltage should be stabilized, and this fraction depends on the ratio of primary to half-secondary voltage which must not be less than unity approximately.

(3) It is advisable to use diodes with a high forward conductance.

(4) The diode loading should be such as to reduce the secondary Q to one-fourth or less of its unloaded value.

(5) When the input signal falls, the secondary Q rises and more series impedance is transferred from the secondary to the primary. This drops the primary voltage but the effect is offset by a reduction in the diode loading of the primary, and by proper choice of circuit parameters the gain to the primary can be made more or less constant. This precaution is necessary to minimize the response of the system to amplitude modulation.

(6) The circuit is somewhat more critical than the conventional phase discriminator circuit in regard to tolerances, the coupling between primary and secondary and the need for balance being particularly important.

(7) The discharge time constant for the stabilizing voltage should be about 0·2 sec. Smaller values permit residual amplitude modulation and larger values lead to difficulties in tuning similar to those experienced with conventional receivers having too long an a.g.c. time constant.

(8) The ratio detector differs from limiter-discriminator circuits in that although the output of the receiver is insensitive to amplitude modulation at audio-frequencies, it does depend on the mean signal level. Automatic-gain control may therefore be found desirable, in which case the control voltage can be obtained from across the condenser C.

The quality of reproduction of an F.-M. receiver depends on the alinement of the discriminator, and correct adjustment is difficult to ensure without elaborate equipment. From the ratio detector it is interesting to turn therefore by way of contrast to the cycle-counting discriminator. This device is relatively insensitive

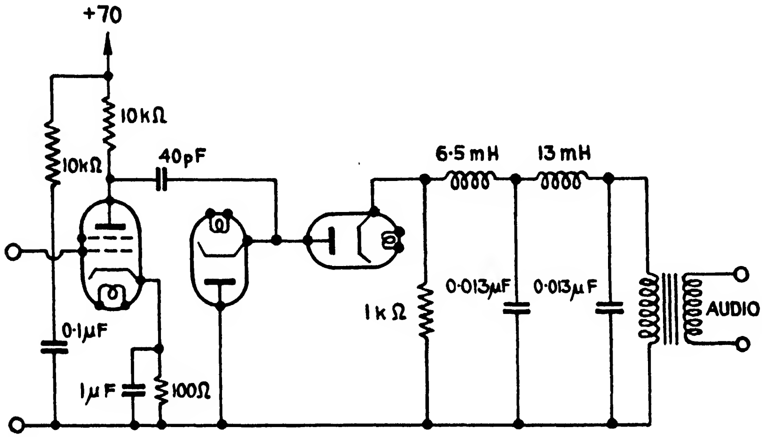


Fig. 9·11. Practical cycle-counting discriminator circuit with values.

but does not involve close tolerances, and risks such as that of distortion due to insufficiently skilled servicing are eliminated as there are no i.f. circuits to be alined; for similar reasons it seems likely to commend itself particularly to the home constructor who has no signal generator at his disposal. A practical circuit* is illustrated in fig. 9·11. The i.f. signals are applied to a limiter which converts them to the form of a square wave of constant amplitude and varying frequency, and the square wave is converted into a succession of uniform short pulses by the differentiating circuit† comprised by the 40 pF. condenser and the 1000-ohm resistance. Differentiation normally produces positive and negative pulses, but the parallel diode removes the negative pulses and the series diode accepts the positive pulses so that flowing through the 1000-ohm resistance we have a succession of short positive pulses, one for each i.f. cycle. Provided the differentiating circuit time constant is short compared with the i.f. period, the pulses are all

* T. Roddam, 'Why align discriminators?' *Wireless World*, July 1948.
† See p. 152.

similar and the mean current is proportional to the intermediate frequency. This mean current varies at the appropriate audio-frequency which is extracted by means of the filter. The circuit is designed to work with a mean input frequency in the region of 150 kc./sec., and the simplest form of receiver employs for its i.f. amplifier a typical video amplifier circuit passing frequencies from, say, 50 to 300 kc./sec.

9·8. Logarithmic receivers

A logarithmic receiver is defined as one having its output proportional to the logarithm of its input. Receivers of this type* have various uses in radar technique, and other possible applications such as automatic field strength recording can also be envisaged.

One method uses the successive detection principle in which diode second detectors are employed after each of the last few i.f. stages. The outputs of the detectors are added up by applying them to buffer amplifiers sharing a common anode load; if the output of any stage increases up to a steady saturation value V_s as the input is increased, the total output voltage with n stages saturated will be nV_s, and the increase from V_s to nV_s clearly requires the receiver input voltage V_i to be increased in the ratio g^{n-1}, where g is the stage gain. If n is not too small we have V_i (approximately) proportional to g^n so that n and therefore V_0 is proportional to $\log_g V_i$. If each stage were linear up to the saturation level the input-output characteristic would be made up of a number of straight sections, but in practice the normal degree of non-linearity is such as to give a fairly smooth curve approximating to the desired law. The required flat overload characteristic may be achieved by juggling with electrode voltages and the use of grid leaks, or excessive overload may be prevented by adjusting the anode loads and electrode voltages so that the overload level of any stage is just sufficient to drive the next stage to the same point. In this latter case the stage gain at saturation is only unity, but the low-level gain will be greater than unity and the circuit loading required is quite likely to coincide with the value desirable from considerations of band-width.

* S. N. Van Voorhis, *Microwave Receivers*, Radiation Laboratory Series, vol. 23, chapter 21, McGraw Hill Book Co. Inc. (1948). R. V. Alred and A. Reiss, 'An anti-clutter radar receiver', *J. Instn Elect. Engrs*, 94, part III (Nov. 1948). See also p. 136.

An alternative system,* of greater simplicity though restricted to somewhat narrower band-widths, employs diodes as non-linear damping resistances across the i.f. circuits. The basic arrangement is shown in fig. 9·12, and fig. 9·13 illustrates the characteristics for various diode bias voltages of an actual amplifier which embodied five such stages. It will be seen that for a bias voltage of −0·2 V. the response was approximately logarithmic for input signals between 4 μV. and 0·4 V. The mean i.f. was 13·5

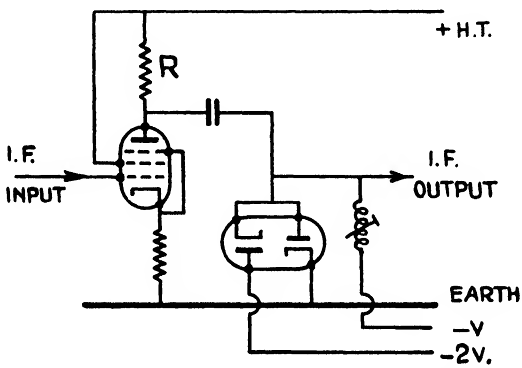


Fig. 9·12. Basic circuit of logarithmic amplifier stage using the damping method.

Mc./sec., and the band-width 2 Mc./sec. with R equal to 3000 ohms. The diodes used (type CV 140) were carefully selected to have similar characteristics, as an alternative to individual adjustment of bias for each stage. The non-linear stages were followed by a linear stage of 20 db. gain, and all the circuits were tuned to the same frequency. Since the band-width of each circuit varies with signal strength the extension of this method to coupled or staggered circuits presents considerable difficulty, and the maximum band-width obtainable is therefore subject to the limitations of 'isochronous' tuned circuits explained on p. 91.

The successive detection type of logarithmic receiver is generally described as a 'linear-logarithmic' receiver. It is to be noted, however, that with *either* of the methods described, the input-output characteristic is linear for small inputs, provided the detector is linear or the output measured before detection. It is only

* R. V. Alred, loc. cit.

when the input is large enough in the one case to overload the final i.f. stage or, in the other case, to cause a significant change in the impedance of the last damping diode, that the characteristic starts to be logarithmic. If the gain is sufficient to bring the noise level well up on to the logarithmic part of the characteristic the

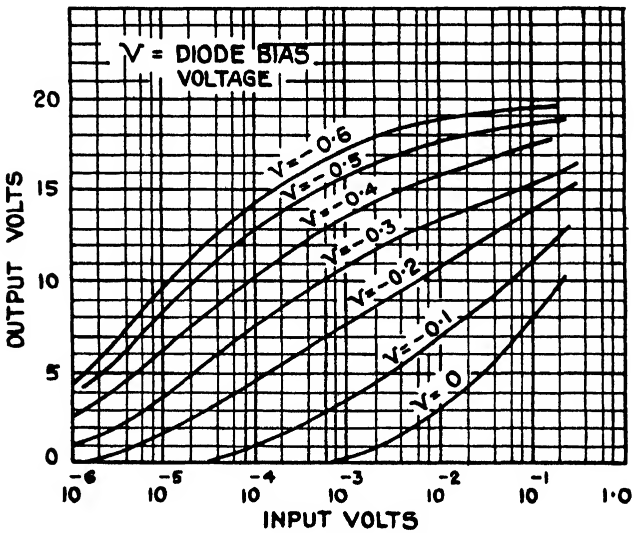


Fig. 9·13. Measured input-output characteristics of damping type of logarithmic i.f. amplifier for various bias voltages.

linear portion has no practical significance, and the receiver or amplifier can be regarded as truly logarithmic; any distinction between 'lin-log' and 'log' receivers should therefore be a matter of gain and noise level rather than of the particular circuit used.

A logarithmic (or 'lin-log') characteristic may be obtained using a linear amplifier in conjunction with a conventional a.g.c. system, by juggling with the valve characteristics and circuit constants. Usually, however, it is difficult to achieve sufficiently rapid response because the a.g.c. time constant must obviously be short compared with the period of any modulation which it may be desired to reproduce logarithmically. In general, this condition leads to instability unless the band-width can be made much wider than the optimum value determined by other considerations such as signal-to-noise ratio.

Appendix

DESIGN FORMULAE FOR INTERMEDIATE FREQUENCY AMPLIFIERS

1. The variation of amplitude and phase with frequency and the time delay, for pairs of similar circuits

Let us first of all consider a single-tuned circuit of the form shown in fig. A1. The impedance Z_s of R, L and C in series is $R+j(\omega L-1/\omega C)$. If for ω we write $(\omega_0+\delta\omega)$, where ω_0 is the resonant frequency (i.e. $\omega_0 LC=1$), and if $\delta\omega$ is small compared with ω_0, we have $Z_s \simeq R+2j\delta\omega L$. The current i, produced by the

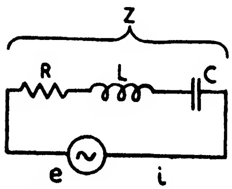


Fig. A1. Single-tuned circuit.

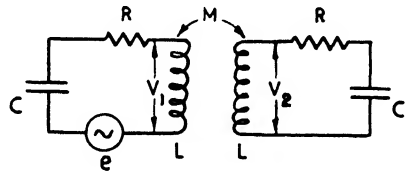


Fig. A2. Coupled circuits with equal loading. M may be replaced by equivalent capacitive or inductive coupling.

voltage e, is equal to e/Z_s, and the voltage measured across L is proportional to i_1. (This is 3 db. down when $2\delta\omega L=R$, and putting $\delta\omega/\pi=B_t$ we obtain the well-known formula for band-width: $B_t=R/2\pi L$.) If the circuit is off-time by an amount aB_t, we have $Z_s=R(1+2ja)$ and the response relative to that at ω_0, expressed as a voltage ratio, is

$$|1/(1+2ja)|=1/\sqrt{(1+4a^2)}. \tag{6·3}$$

Now let us take an exactly similar circuit coupled to the first by a mutual inductance M, and calculate the voltage V_2 across the secondary circuit resulting from the voltage e in series with the primary. Referring to fig. A2 we may equate the applied voltage to the voltage drop for each circuit, so that for the primary

$$e+j\omega M i_2 \simeq i_1 R(1+2ja)$$

and for the secondary

$$j\omega M i_1 \simeq i_2 R(1+2ja).$$

Solving for i_2 and substituting the arbitrary condition $kR=\omega M$, we have

$$V_2/e=j\omega Li_2/e=\frac{kj\omega L/R}{1+k^2-4a^2+4ja}, \tag{6·4}$$

which compares with $V_1/e=j\omega L/R$ for a single-tuned circuit and has a maximum value of $\omega L/2R$ obtained when $k=1$. This is the well-known condition for critical (or transitional) coupling, and k may be defined as the 'Number of times critical coupling'.

The shape of the curve of relative amplitude A versus frequency is thus given by

$$A\approx\frac{1+k^2}{\sqrt{[(1+k^2-4a^2)^2+16a^2]}} \tag{6·5}$$

and the phase response by

$$\tan\phi\approx\frac{4a}{1+k^2-4a^2}. \tag{6·6}$$

The notation used is convenient, because it enables response curves to be derived in a general form and then applied to any specific case simply by putting $a=\delta\omega/2\pi B_t$. The foregoing analysis still holds approximately (in all practical cases) for unequal values of primary and secondary inductance L_p, L_s, provided the circuits have equal Q, and L (equation 6·4) is replaced by $\sqrt{L_pL_s}$.

Let us next see what happens with circuits in cascade, not coupled to each other, staggered in tuning symmetrically with respect to ω_0, and separated by a frequency difference pB. The relative response A at any frequency $\omega_0/2\pi\pm aB$ is given as follows by a simple extension of the reasoning for a single circuit:

$$A=\frac{(1+jp)(1-jp)}{[1+2j(a-\frac{1}{2}p)][1+2j(a+\frac{1}{2}p)]}=\frac{1+p^2}{1+p^2-4a^2+4ja}. \tag{6·7}$$

This is identical with equation (6·5) except for the substitution of p for k, and similarly the phase response is given by equation (6·6) replacing k by p. Substituting $p/2$ for a in equation (6·3) we find that the mid-band gain per stage is $1/\sqrt{(1+p^2)}$ times the on-tune gain $g_m/2\pi CB_t$ of a single circuit.

2. Coupled pairs with loading on one side only

If the band-widths of a pair of staggered circuits are unequal, it is obvious that the response curve must be unsymmetrical, whereas in the coupled circuit case there is no disturbance of symmetry by unequal loadings. Let us examine the simple case when all the secondary loading in fig. A2 is transferred to the primary, or vice versa, there being no other physical change, so that the equivalent series resistances for the two circuits are respectively $2R$ and zero. The definition of a and k in terms of R, L and M will not be altered except that it will be convenient to distinguish between this and the previous case by writing k_u instead of k.

Proceeding as before, we have

$$\frac{V_2}{e}=\frac{jk_u\omega L/R}{k_u^2-4a^2+4ja}, \tag{6·8}$$

$$A=\frac{k_u^2}{k_u^2-4a^2+4ja} \tag{6·9}$$

and
$$\tan\phi=4a/(k_u^2-4a^2). \tag{6·10}$$

For A and ϕ to be the same in each case, we require

$$k_u^2=1+k^2. \tag{6·11}$$

From equations (6·8) and (6·4) the gain in the case of unilateral loading is greater than the gain with equal loading by the factor $(1+k^2)/kk_u$, i.e. for similar frequency response the gain improvement factor is

$$\sqrt{(1+1/k^2)}. \tag{6·12}$$

With transitional coupling the improvement is $\sqrt{2}:1$ but reduces to 1·19 to 1 for a 50% increase of k.

The response is now much more critical to coupling factor; thus for transitional coupling k_u is $\sqrt{2}$, and reducing it to unity has a similar effect on response to that of reducing k from 1·0 to zero.

If k_u is zero, the response is infinitely sharp, since we have assumed one of the circuits to have zero loading.

3. Calculation of '3 db. down' band-width for pairs of circuits

The band-width of a pair of circuits between points at which the amplitude is A_r times the amplitude at the centre frequency may be found from equations (6·5), (6·7) or (6·9) by putting $A=A_r$ and solving for a. From equation (6·5) we obtain

$$4a^2=k^2-1+\sqrt{\left[\left(\frac{1+k^2}{A_r}\right)^2-4k^2\right]}. \tag{6·13}$$

Starting from equation (6·7), k in equation (6·13) is replaced by p.

To find the band-width of n similar pairs of circuits to 3 db. down, we put $A_r = 1/\sqrt[n]{2}$ and evaluate $2a$ from equation (6·13). The required band-width is then given by $2aB_t$.

4. The amplitude of humps produced by over-coupling or over-staggering

Differentiating equation (6·5) with respect to a, and equating to zero, we find that the amplitude of response is a maximum when $2a = \sqrt{(k^2-1)}$, which is real provided k is greater than 1.

By putting $a=0$ and $a=\frac{1}{2}\sqrt{(k^2-1)}$ in equation (6·5), we find that with over-critical coupling,

$$\frac{\text{maximum amplitude of response}}{\text{amplitude at mid-frequency}} = \frac{1+k^2}{2k}. \tag{6·14}$$

For over-staggering, of course, p replaces k.

5. Time delay

As explained in § 6·4·2, the time delay τ is given by $d\phi/d\omega$. From equation (6·6), we have

$$\phi = \tan^{-1}\frac{4a}{1+k^2-4a^2} \quad (\text{where } a = \delta\omega/2\pi B_t)$$

for a coupled pair of similar circuits, so that for each pair

$$\tau = \frac{2}{1+k^2}\frac{1}{\pi B_t},$$

and for m pairs

$$\tau = \frac{2m}{\pi B_t(1+k^2)}. \tag{6·15}$$

With $m=1$ and critical coupling this simplifies to $\tau = 1/\pi B_t$ and is the same as the delay time ($d/d\omega \tan^{-1}2a$) for one circuit alone.

6. Response when the ratio $\omega_0/2\pi B_t$ is small

In the cases so far considered it does not matter whether the circuits are damped by a series resistance R_s or a parallel resistance R_p, subject to the well-known equivalence $R_s = 1/\omega^2C^2R_p$. Although this can only be true for one frequency at a time, it holds throughout the receiver band-width to a sufficient degree of approximation provided $2\pi B_t \ll \omega_0$.

If we now abandon this condition, the effect on the response shape depends on whether the damping is by means of series or parallel resistances. Let us take first the single-tuned circuit of fig. A3, and let Z_p be the resultant of R_p, C and L in parallel. If this circuit comprises an interstage coupling, the gain is proportional to Z_p or $1/(1/R_p+1/j\omega L+j\omega C)$ and the ratio of gain at frequency $\omega/2\pi$ to gain at the resonant frequency $\omega_0/2\pi$ is therefore

$$j\omega L/[R_p(1-\omega^2LC)+j\omega L].$$

Writing $\omega_0^2LC=1$ this reduces to

$$\frac{1}{1-jR_p\omega_0C(\omega_0/\omega-\omega/\omega_0)}. \qquad (6\cdot16)$$

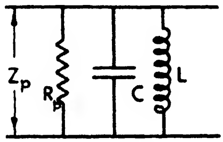


Fig. A3. Tuned circuit with parallel damping.

From inspection of this it is evident that the frequency response has geometric symmetry with respect to the resonant frequency, the response at frequencies $\omega_1/2\pi$ and $\omega_2/2\pi$ being equal provided $\omega_1/\omega_0=\omega_0/\omega_2$.

Considering now a staggered pair of circuits of the fig. A3 type, tuned to ω_a and ω_b; their responses are respectively proportional to

$$\left|\frac{1}{1-j\,R_{pa}\omega_a\,C(\omega_a/\omega-\omega/\omega_a)}\right|$$

and

$$\left|\frac{1}{1-j\,R_{pb}\omega_b\,C(\omega_b/\omega-\omega/\omega_b)}\right|,$$

where R_{pa} and R_{pb} are the values of R_p at frequencies $\omega_a/2\pi$ and $\omega_b/2\pi$ respectively.

The combined response has geometric symmetry with respect to ω_0 provided $\omega_a/\omega_0=\omega_0/\omega_b$ and provided *also* that

$$R_{pa}\omega_aC=R_{pb}\omega_bC,$$

i.e. that the circuits have the same power factors at resonance.

For the case of series damping, R_p is replaced by $1/\omega^2C^2R_s$ and calculation becomes much more involved. We need not interest ourselves in this case, since it is usually at least equally convenient to use parallel damping.

In the coupled circuit case, symmetry depends on the variation with frequency of the power factor of the individual circuits. If the series resistance consists of coil losses it is a function of frequency, and it generally happens that the power factor F is approximately constant. Following the procedure of § 1, avoiding the approximations made there, but assuming that F is constant, we obtain expressions similar to equation (6·4), (6·5), etc., with a replaced by $a' = (\omega^2 LC - 1)/2F$, a' and a becoming the same of course when $\delta\omega/\omega$ is small. The response is again symmetrical with respect to ω_0 for frequencies $\omega_1/2\pi$, $\omega_2/2\pi$, such that

$$\omega_1/\omega_0 = \omega_0/\omega_2.$$

7. Triple-staggered circuits

A close approximation to a flat-topped response may be obtained by combining the response of a single-tuned circuit with that of a pair of over-coupled or over-staggered circuits. To obtain the response of each group of three circuits the single-circuit response may be multiplied by that of the pair, but as the circuits have different band-widths the value of a appropriate to the single circuit is not the same as for the pair. Referring to equation (6·3), we will write the response of the single circuit as $1/\sqrt{(1+4n^2a^2)}$, defining a with respect to the pairs, so that combining equations (6·3) and (6·7),

$$A = \frac{1+p^2}{\sqrt{\{[1+4n^2a^2][(1+p^2-4a^2)^2+16a^2]\}}}. \tag{6·17}$$

There is considerable latitude in the amount of staggering provided the single circuit is carefully fitted to the dip in the response of the staggered (or coupled) pair. If the single circuit is too broad, the hole is not fully filled in and the curve remains double-peaked, and if it is too sharp there will be a centre peak of relatively large amplitude. The mid-band gain is evaluated most easily by treating each stage separately in accordance with § 1.

A simple method of design is to plot families of single circuit and over-staggered pair response curves such as those of fig. A 4 which have been calculated from equations (6·3) and (6·7). These curves, preferably re-drawn on a larger scale, may be traced, held up to the light, and fitted together by trial and error so that a single-circuit response curve, upside down, fits closely into the dip between the peaks of the other.

The following relationships may be found useful; let us first try making the response at the centre frequency equal to that at the

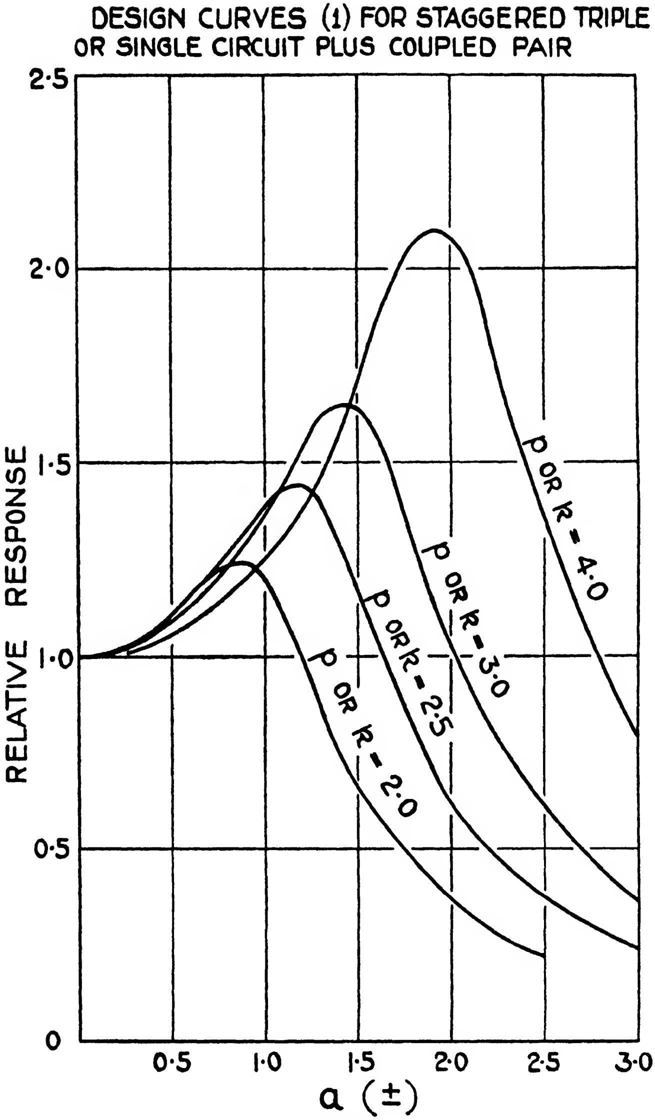

DESIGN CURVES (1) FOR STAGGERED TRIPLE
OR SINGLE CIRCUIT PLUS COUPLED PAIR
2·5
2·0
1·5
1·0
0·5
0
RELATIVE RESPONSE
0·5
1·0
1·5
2·0
2·5
3·0
a (±)
p OR k = 4·0
p OR k = 3·0
p OR k = 2·5
p OR k = 2·0

Fig. A4(*a*).

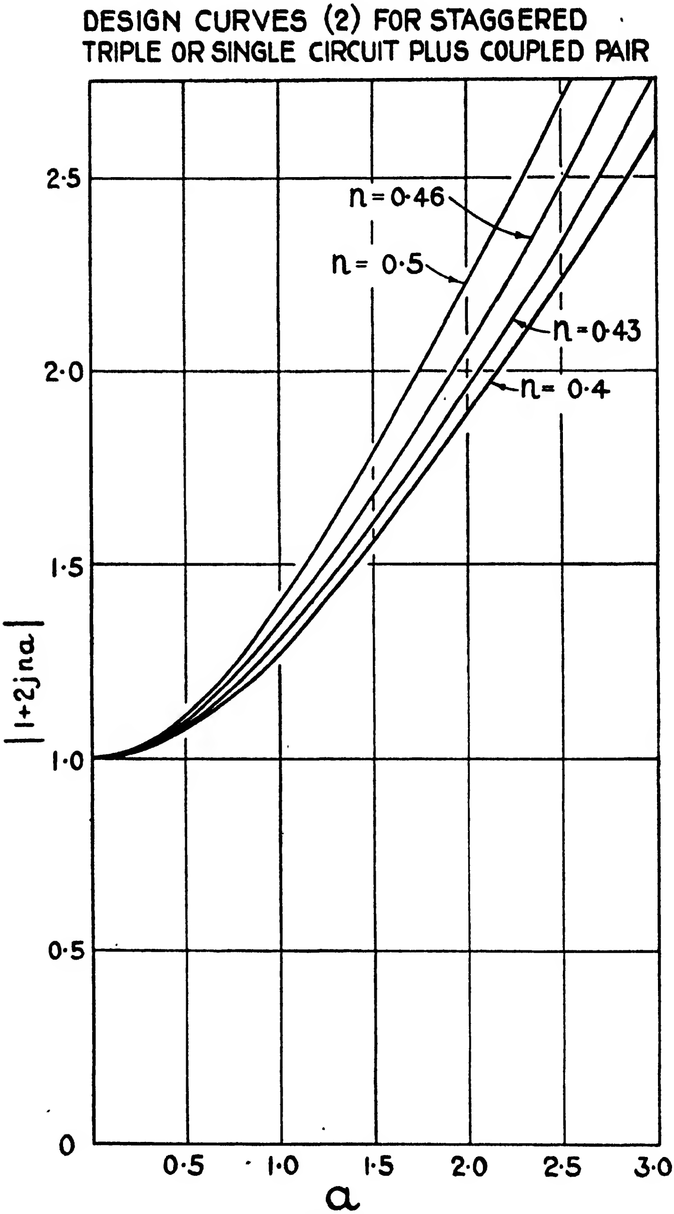

DESIGN CURVES (2) FOR STAGGERED
TRIPLE OR SINGLE CIRCUIT PLUS COUPLED PAIR
n = 0·46
n = 0·5
n = 0·43
n = 0·4
2·5
2·0
1·5
1·0
0·5
0
|1+2jna|
0·5
1·0
1·5
2·0
2·5
3·0
a

Fig. A4(*b*)

frequencies of the humps of the outer pair. We can see from equation 6·14 that this requires

$$\left.\begin{aligned} 1/\sqrt{(1+4n^2a^2)} &= 2p/(1+p^2) \\ \text{and} \qquad 2a &= \sqrt{(p^2-1)}, \\ \text{whence} \qquad n &= \sqrt{(p^2-1)}/2p. \end{aligned}\right\} \qquad (6\cdot 18)$$

On plotting the corresponding response we find, if the amount of staggering is large, that the peaks have been brought in towards the centre frequency and not entirely removed, so that a further approximation taking the form of a reduction of centre circuit band-width may be necessary.

Another relationship which may be of interest is obtained by equating the centre frequency curvature of the single circuit response to minus that of the staggered pair, so that the overall response at the centre frequency is flat. This gives

$$n = \frac{1\cdot 4\sqrt{(p^2-1)}}{p^2+1}. \qquad (6\cdot 19)$$

There is rough agreement between equations (6·18) and (6·19) for values of p ranging from about $1\frac{1}{2}$ to 3, but in general equation (6·18) gives a closer approximation to the required answer.

8. Maximum gain of amplifier having single circuits all tuned to the same frequency

Let the amplifier consist of q stages. From equation (6·3) the relative overall amplitude response is given by $1/(1+4a^2)^{\frac{1}{2}q}$, and to find the band-width to 3 db. down we have to equate this to $1/\sqrt{2}$. For a large number of stages qa^2 will be small and we can therefore write

$$1+2qa^2 \simeq \sqrt{2}, \quad \text{i.e.} \quad 2a \simeq 1/1\cdot 1\sqrt{q}.$$

The band-width of each individual circuit must therefore be $1\cdot 1\sqrt{n}$ times greater than in the case of a single-stage amplifier of similar overall band-width, and the stage gain is reduced in this ratio.

The overall gain is therefore given by

$$\left(\frac{g_m}{2\cdot 2\pi C_v B_0}\right)^q \left(\frac{1}{q}\right)^{\frac{1}{2}q}. \qquad (6\cdot 20)$$

As q is varied, this has a maximum value when

$$q = \frac{1}{\epsilon}\left(\frac{g_m}{2 \cdot 2\pi C_v B_0}\right)^2, \tag{6·21}$$

C_v is the total tuning capacity and B_0 the overall band-width. Any increase in the number of stages above the nearest whole number to q leads to a *reduction* in gain, assuming the band-width to be held constant.

INDEX