RADIO COMMUNICATION AT ULTRA HIGH FREQUENCY

RADIO COMMUNICATION AT ULTRA HIGH FREQUENCY

BY

JOHN THOMSON

M.A., D.Sc., F.Inst.P.

Professor of Physics and Electrical Engineering in the Royal Naval College, Greenwich

WITH 3 PLATES AND 85 DIAGRAMS

METHUEN & CO. LTD. LONDON

36 *Essex Street, Strand, W.C.2*

First published in 1950

CATALOGUE NO. 5189/U

PRINTED IN GREAT BRITAIN

PREFACE

THE aim of this book is to provide an account of modern developments in telecommunications employing radio waves of lengths ranging from a few metres to a few millimetres. Since World War II much has been written round the use of this part of the spectrum for radar, but little has been published in book form regarding the corresponding advances which have taken place in radio communication. The potentialities of the new ultra high frequency techniques are fascinating, and in the United States of America have already reached some measure of fulfilment. Both sound and vision broadcasting may radically alter their character as a result of these developments: already mobile radio services such as those required by the police and by coastal shipping have become practical propositions: while point-to-point multiplex telephony, television relays, and colour television are even now contesting for the use of the ultra high frequency bands.

The present volume is written in language which, I hope, will be intelligible both to the radio student and to the practising communication engineer, but the devices described may also be of value to workers in other electronic fields. For that reason a fairly detailed treatment of certain velocity modulation valves has been given.

No attempt has been made to obtain that uniformity of diagrammatic representation and symbolism which is so necessary in detailed engineering. Although I have the greatest respect for such standardization, and although its logical application is "a consummation devoutly to be wish'd," the student and engineer should have no trouble in following the more haphazard symbolism with which many of us have lived for many years. Similarly, no particular attention has been paid to units. Once more I heartily agree that it would be a good thing if we all used the rationalized M.K.S. system, and once more I feel that my readers will be in no real difficulty even where I mix electrostatic and electromagnetic units in one equation.

The book is largely written round the war and post-war work of my recent colleagues and myself in what is now Royal Naval Scientific Service. To these colleagues my best thanks are due; and to the Board of Admiralty for permission to publish the work I also wish to record my gratitude. The views expressed, however, are entirely my own responsibility.

The chapter on Receiver Input Circuits which constitutes the major contribution on reception has been written by my friend and recent colleague, Mr. Peter Trier, M.A. My debt to him for this and for much useful discussion of the subject is very great.

For the Plates I am indebted to several sources. Figures 1 and 2 of Plate II were kindly supplied by the makers of the valves, the Research Laboratories of the General Electric Company at Wembley. Figures (*a*), (*b*), (*c*), and (*e*) of Plate III were given by the makers of the crystal units, the Marconi Wireless Telegraphy Company, Chelmsford. For permission to publish Figures 1, 2, and 4 of Plate II am indebted to the Board of Admiralty.

To my wife for her care with the manuscript and the Index my best thanks are also due. Without her help and encouragement the work would not have been completed.

J. THOMSON

CONTENTS

V. MODULATION TECHNIQUES

VI. FREQUENCY CONTROL

VII. COMMUNICATION SYSTEMS

PLATES

INTRODUCTION

What is meant by the composite term "ultra high frequency"? The name is clearly unscientific and would never find a place in a rational nomenclature. It has been used in the title of the present work simply because there is no better term in common use, and because the domain in which we are interested has boundaries which are almost as vague as the name itself. Historically, the science of radio communication has advanced steadily towards higher and yet higher frequencies. Each generation of radio engineers, amateur and professional, has imagined that its use of the highest frequency available in its day represented a rather extraordinary achievement. It was in this way that the term high frequency came to be used. The particular generation of radio enthusiasts which first employed frequencies of the order of a million cycles per second felt that this was a high frequency compared with those used by Marconi and his *confreres*. The next generation feeling its way into the hundred megacycle region was forced to employ the term "very high frequency" to denote the change. Then, when frequencies of the order of a thousand megacycles per second began to be used, the pioneers in the new part of the spectrum had to invent a term which suggested something higher than "very high." "Ultra high" was the outcome of this terminological tangle. On parallel lines, speaking of wave-length, but in a somewhat different part of the spectrum, the terms "short waves" and "ultra short waves" were introduced.

A scientific nomenclature must be based on the decimal system, and appropriate words must be invented to describe the different orders. But with frequencies the values of which run to one hundred thousand million cycles, there is a certain difficulty. It is probably simpler, and just as scientific, to speak in terms of wave-length (as in the early days), and if we do so, a perfectly satisfactory decimal centimetric nomenclature suggests itself. The following table illustrates this suggested system.

Frequency Band	Wave-length Band	Name
3.10^4 to 3.10^5 cycles/sec.	10 to 1 kilometres	Kilometric waves
3.10^5 to 3.10^6 cycles/sec.	10 to 1 hectometres	Hectometric waves
3.10^6 to 3.10^7 cycles/sec.	10 to 1 decametres	Decametric waves
3.10^7 to 3.10^8 cycles/sec.	10 to 1 metres	Metric waves
3.10^8 to 3.10^9 cycles/sec.	10 to 1 decimetres	Decimetric waves
3.10^9 to 3.10^{10} cycles/sec.	10 to 1 centimetres	Centimetric waves
3.10^{10} to 3.10^{11} cycles/sec.	10 to 1 millimetres	Millimetric waves

In this book we shall be concerned with radio communication techniques and systems applicable at frequencies greater than 100 Mc/s. It will be seen, therefore, that, using the above nomenclature, the subject-matter relates to metric, decimetric, centimetric and millimetric wave-lengths.

Historical Background. It is only recently that use has been made of the radio frequencies above 100 Mc/s, but it is of some interest that the original experiments of Hertz[1] employed electrical resonators the frequencies of which were in the 70 Mc/s region. Indeed some of Hertz' later experiments, using reflectors to focus the electromagnetic radiation, employed frequencies in excess of 100 Mc/s, and old apparatus, made in Lord Kelvin's time, and until recently retained in the museum of the Natural Philosophy Institute of the University of Glasgow, was suitable for use with decimetric waves.

During the fifty years succeeding Hertz' experiments the interest of radio engineers, originally focused on the "medium frequencies," gradually moved towards shorter and shorter wave-lengths, until, not long before Great War II, thermionic valves had reached a stage of development at which the production of easily measurable powers at the "ultra high frequencies" had become possible. Although it was not generally known at that time, the development in Great Britain of "radio-location" had given a certain impetus to the search for better ultra short-wave thermionic valves, and the General Electric Company had in this connection been highly successful.[2] At the same time and quite independently, vacuum physicists in Japan,[3] America[4] and Britain[5] had suggested such improvements to the original idea of the "magnetron oscillator" that devices were coming on to the market capable of producing continuous oscillations at frequencies up to about 1500 Mc/s. It was about the same period (1936) that the principle of the "velocity modulation" of an electron beam was enunciated,[6] and not long after this the first so-called "klystron" oscillators, employing the velocity-modulation effect, were made in America.[7] These new oscillators were particularly well adapted to the production of oscillations of decimetric wave-length, and so a new tool was made available for the exploration of the ultra high frequency region.

Nor did the spate of discoveries cease at this point. In America Barrow, Southworth, and others[8] were working out the theory of the now familiar "wave-guide," and demonstrating, experimentally, that a metal tube formed an almost perfectly non-attenuating guide for waves of high enough frequency.

Probably the first British experiments, employing the newly discovered techniques, were made by H. R. L. Lamont in the

University of Glasgow in the years 1937–8–9,[9] and, considering the importance of such devices in the later stages of World War II, it is interesting to note that Lamont's first guide consisted of a piece of boiler tube salvaged from one of the scuttled German battleships of World War I. It was five feet long and seven inches in diameter. With it Lamont employed a G.E.C. magnetron, using this valve at wave-lengths down to 15 cm.

As was to be expected, the beginning of the war saw enormous strides in decimetre wave techniques. First in importance was the development by Randall and his colleagues at Birmingham University of the multi-resonator magnetron. This provided the transmitting valve for decimetric radar, and was almost immediately followed by the reflection klystron, developed by Admiralty Signal Establishment Group at Bristol from the theory already expounded by the Varian brothers[7] and by Hahn and Metcalf.[10] This reflex klystron provided the oscillator valve for the decimetric radar receiver. Thereafter, development proceeded at a phenomenal pace. One of the most important elements in the new radar equipment was the receiver frequency changer or first detector. For this the tungsten-silicon rectifier was found to be very efficient, and concentrated research succeeded in improving the performance of this component in the most amazing fashion. Some idea of the advance may be obtained from the following figures. The first silicon crystal used in radar experiments had a noise factor of about 25 db. Towards the end of the war crystals used at the same frequency were discarded if their noise factor was greater than 9 db.

There is, however, another and less favourable aspect of the war-time advances. All available effort was directed towards the detection and location of targets, and the orthodox radio communication possibilities of these ultra short waves were largely neglected. There was one outstanding exception in the development of the No. 10 Set for Army use,[11] but, in general, communication techniques attracted little attention. Indeed, looking back over war-time radio history, there were many occasions when the exigencies of the radar problems led to the invention of devices which inhibited the growth of healthy communication techniques. In some measure radio communications at ultra high frequencies are now suffering from the very one-sided war-time programme.

Despite these limitations important advances have been made. In particular, thermionic valves suitable for the particular tasks of communication have been developed, and of these the disc-seal triode is the most important. Techniques have been worked out for the stabilization of the frequency of transmitters and receivers, so that present-day stability at 3000 Mc/s compares favourably with pre-war stability at 3 Mc/s. The problem of the modulation of an ultra high frequency carrier has been attacked with considerable success. Aerials suitable for communication have been developed at all frequencies, and last (but not least), a very large mass of accurate information is now available on the propagation of metric, decimetric and centimetric waves.

Although it is true that ultra-short wave communication is only in its infancy, enough research has already been done to place the subject on a secure and promising foundation. Many misconceptions and prejudices have been swept away, and sufficient information has been gleaned to ensure the technical success of ambitious communication projects. It is too early to say how important the new radio spectrum will be in the future, but, judging by the uses to which it has already been put in the communication networks of the world, there is a well defined demand for ultra high frequency engineering.

The Characteristics of Ultra High Frequency Radio Systems. There are many factors which limit the utility of radio links which employ frequencies below 100 Mc/s. Some of the more important of these are as follows—

(*a*) Atmospheric and man-made static.

(*b*) The impossibility of containing the energy inside a given area.

(*c*) The necessity for large aerial structures and for the use of large transmitting powers.

(*d*) The impossibility of employing "wide-band modulation."

(*a*) To exemplify factor (*a*) it is only necessary to call attention to the fact that it is useless to manufacture the most sensitive receiver which can be designed in the medium frequency and high frequency regions of the radio spectrum, since the electrical noise entering the receiver from its aerial is many times greater than the "receiver" noise itself. For that reason the signal to noise ratio achieved in a medium or high frequency communication link is usually much less than one would expect it to be in the absence of static. The frequencies above 100 Mc/s are surprisingly free from static. Not only are atmospheric noises almost completely absent, including the well-known but erroneously named "precipitation static," but man-made static is not very troublesome. Ignition noise can still be heard, even at 1000 Mc/s, for the resonant circuits of many commercial spark discharges are of very high frequency; but if steps are taken in the receiver to screen the amplifying sections which operate at lower frequencies, few if any of the usual electrical pulses possess the necessary very high frequency harmonic content to be detected. Hence ultra high frequency receivers are in general only limited in their sensitivity by the noise arising from random electronic movements in their components. As this is an unwanted effect which is the subject of close study, "receiver noise" has a continually decreasing value, and the signal to noise ratio achieved in ultra high frequency radio links is tending rapidly towards the maximum theoretically possible.

(*b*) Factor (*b*) is a more controversial one. No radio frequency yet employed is such that it can *only* be detected within a certain radius from the transmitter. Propagation peculiarities make it possible for even the radiations of shortest known wave-length to be bent round the earth in such a fashion that they can be detected far beyond their "normal" range. But from a practical point of view it is unlikely for radiation of very high or ultra high frequency to be picked up much beyond the "line of sight" distance from the transmitting aerial. These radiations do not easily bend round the earth, and so it is possible in practice to assume that the energy from an ultra high frequency transmitter does not appear outside a certain area of the earth's surface. This fact is, of course, of fundamental value for certain types of radio communication services.

(*c*) A communications aerial, be it transmitting or receiving, is the most efficient disseminator or collector of energy when its length is of the same order as the wave-length of the radiation concerned. For this reason the aerial length appropriate to the use of hectometric waves is of the order of hundreds of metres. This is a large (and usually heavy) structure, and the need for it sets a severe practical limitation upon the use of hectometric waves. If, indeed, we wish to direct such waves towards one point of the compass and for that reason wish to erect a directional array of aerials, the size and cost of the structure becomes almost prohibitive.

At ultra high frequencies, the same principle is operative. Again radiators and collectors should have the same order of linear dimension as the wave-length at which they are used. Since these wave-lengths may be very small, extremely efficient aerials can be achieved, and directional arrays are possible at reasonable size and cost. To illustrate what can be done it need only be mentioned that a three-foot diameter paraboloid, used at a wave-length of 10 cm. has a gain over a simple dipole of the order of two hundred.

(*d*) Modern communication systems include many links across which it is desired to transmit a very large quantity of information in the minimum time. The most outstanding example of this is the multi-channel system in which a number of independent two-way conversations may be carried on simultaneously. But by a well-known theorem in radio engineering, the frequency band containing the modulation energy must increase in width as the rate of transfer of information increases. For example, all the intelligence in a normal conversation can be transferred by a block of frequencies about 3000 c/s wide, whereas the transfer of a "live" picture (television) requires a block of frequencies several megacycles per second wide. It is difficult to obtain such wide-band modulation at carrier frequencies of the same order of magnitude, and so it becomes necessary, where the rapid transfer of the maximum amount of intelligence is required, to use very high or ultra high frequencies, where a modulation band-width of the order of a few megacycles per second is small compared with the carrier frequency.

The preceding description of the general characteristics of ultra high frequency radiation immediately suggests the uses to which it may be put in communications. Wherever a beam of radiation is required, as in point-to-point working, the shorter wave-lengths are preferable, and this is probably the most important use to which they can be put. For example, a radio link in a telephone network, spanning an estuary or a narrow sea, will profitably employ a multiplex transmission directed from one land-line terminal to another. For this purpose decimetric or centimetric waves are ideal. For longer links across land, such as television relays, the same is true, assuming for the moment, that the ultra high frequency components have a reasonable "life" and that the maintenance problem is solved. One can imagine a chain of relay stations linking London and Edinburgh and providing thereby many communication facilities such as television at all intermediate points.

The other use of the shorter waves is for mobile radio telephony. The best example is the police radio network. In this case use is made of the limited range of the radiation. The country is divided into districts, each having its own self-contained system, and it is expected that interference from neighbouring districts will be mitigated by the "line of sight" property of short-wave propagation. Such a mobile network must be in some measure a compromise. If a frequency were used which gave true "line of sight" working, operations would be confined to rural districts, where a transmitter could be sited at a considerable height. In practice frequencies high enough to be fairly rapidly attenuated beyond the line of sight are used, but at the same time the frequency is low enough to ensure that the obstacles presented by an urban area are not sufficient to attenuate the radiation beyond a useful level. A good compromise seems to be found in frequencies lying between 70 and 150 Mc/s. A second example of mobile communications is the radio telephony link between aircraft and airport, and in this case the choice of a very high frequency is dictated by more than one factor. "Line of sight" between an aircraft and a ground station increases rapidly as the height of the aircraft increases, and ranges of the order of two hundred miles are possible. Nevertheless, it is useful to employ a frequency such that it is unlikely to interfere with other aircraft at even greater distances, and for that reason alone a very high frequency is preferred. A second factor here of paramount importance is the absence of static in this region of the spectrum, making the link fairly independent of rapidly varying meteorological conditions. The third factor is the radiation efficiency of an aircraft aerial. This is, of course, greater at the higher frequencies. Finally, the very high frequencies are about to be used extensively by the merchant navy for ship to shore telephony and for ship to ship signalling at short distances. The navigation of crowded estuaries such as the Thames and the Mersey is much simplified if control can be exercised by radio telephony. For this purpose the frequencies between 100 and 200 Mc/s are eminently suitable, and are already being used in the United States of America.

Looking into the future, it is not difficult to imagine a host of other uses. Schemes have already been put forward for completely automatic radio-telephony networks, serving subscribers in areas which cannot be "wired," and employing decimetric waves. Communication between railway engine drivers and traffic controllers is already in operation in U.S.A., using centimetric waves. In fact, wherever communication facilities are required, but metallic connections cannot be made (or would be uneconomic) the ultra high frequencies have a part to play. The disabilities of the so-called "medium" and "high" frequencies in the past have prevented the general use of radio communications. That day is past.

Special Problems in Ultra High Frequency Communications. Despite what has been written above, it must not be imagined that there are no difficulties to be overcome in the exploitation of the shorter waves. As in all forms of radio signalling, the propagation of these very high frequencies across the earth's surface is affected by a variety of meteorological conditions which give rise to deviations from the "normal" of considerable magnitude. The subject will be considered later, but it is as well to note at the outset that the normal "line of sight" characteristic of ultra high frequency radiation can be so masked by variations in the refractive index of the troposphere that almost unattenuated propagation over distances up to one hundred miles parallel to the surface of the earth is not uncommon. This so-called "duct" effect may be classed scientifically as "abnormal," but in certain latitudes and at certain times of the year, it is the normal state of affairs. Conversely, in polar regions in particular, it is possible to have an atmospheric condition which limits the range of these radiations to something much less than "line of sight." This phenomenon is less common than the other in habitable regions, but it must not be omitted from calculations, when the possibility of "fading" is of importance.

At one time it was thought that the presence of snow or rain produced serious attenuation at centimetric wave-lengths, and there is no doubt that such attenuation has been observed as a function of the transmitter—propagation path—receiver. It has since been shown, however, that rain or snow produces quite small effects at frequencies below 10,000 Mc/s, and it is probable that the earlier reports signified avoidable losses in aerials and feeder cables.

On the engineering side there are many unsolved problems of which the more outstanding may be mentioned. First, and probably most important of all, the thermionic valves available at decimetric and centimetric wave-lengths are still far from being satisfactory. This is not a serious limitation on development, for it is known that the present valves can be enormously improved. The unsatisfactory state of the art is due in large measure to the emphasis placed upon valves for radar during the recent War. It is to be hoped that enough attention will be given to communications in the future to ensure that the theoretical performance is more nearly attained. At the present moment there is no satisfactory small signal amplifier available for frequencies above 1000 Mc/s, and this may be a genuine limitation due to the nature of thermionic emission. But when it is said that there is a lack of continuous-wave power amplifiers and power oscillators, this only means that the incentive to develop such devices has not been present in the past.

Sooner or later the inertia of the electron will put a limit to the invention of radio devices at higher frequencies. Transit-time effects, so much discussed in pre-War days, have been in some measure overcome as a result of study, and the use of the velocity-modulation principle has removed their menace to even shorter wave-lengths. But the process cannot continue indefinitely, and it seems that the use of electron streams to generate and control electric oscillations must cease to be efficient at some frequency in the neighbourhood of 100,000 Mc/s. At this point macroscopic merges into microscopic, and the realm of ultra high frequency radio comes to an end. The second important problem in the short-wave region is the development of cheap circuits. During the War many investigators and engineers became accustomed to the use of silver-plated components and forgot that such units are doubly expensive in material and workmanship. In this respect a fundamental problem remains, the solution of which is certain but is not yet known. The tuning of such units to cover a wide frequency band presents a problem to which an elegant answer is not yet forthcoming, while the derivation of a standard of frequency suitable for centimetric waves is a costly and elaborate piece of engineering out of all proportion to the value of the result. In short, although most problems in ultra high frequency communication can be solved in the laboratory, the engineer has an arduous road to travel before full advantage can be taken of the new techniques.

REFERENCES

(1) "Electric Waves," Macmillan, 1900.
(2) Gavin, *Wireless Engineer*, 16, p. 287 (1939).
(3) Okabe, *Proc. I.R.E.*, 27, p. 24 (1939).
Uda, *Electrot. J. Tokyo*, 2, p. 291 (1938).
(4) Linder, *Proc. I.R.E.*, 24, p. 633 (1936).
(5) Megaw, *G.E.C. J.*, 7, p. 94 (1936).
(6) Arsenjawa-Heil and Heil, *Zeits. f. Phys.*, 95, Nov. and Dec. 1935.
Bruche and Rechnagel, *Zeits. f. Phys.*, 108, March 1938.
(7) Varian and Varian, *Journ. of Applied Physics*, 10, May 1939.
(8) Barrow, *Proc. I.R.E.*, 24, p. 1298 (1936).
Southworth, *Bell Syst. Tech. J.*, 15, p. 284 (1936).
Schelkunoff, *Proc. I.R.E.*, 25, p. 1457 (1937).
(9) *Phil. Mag.*, 29, p. 521 and 30, p. 1 (1940).
(10) *Proc. I.R.E.*, 27, p. 106 (1939).
(11) *J.I.E.E*, 94, IIIA, No. 11, p. 105 (1947).

CHAPTER I

CIRCUITS FOR USE AT ULTRA HIGH FREQUENCIES

1.1. **Frequency Characteristics of Passive Circuit Elements.** The practical realization of a component possessing approximately only one of the properties of inductance, resistance or capacitance is a matter of some difficulty even at moderately high frequencies. At ultra high frequencies the difficulty becomes exaggerated. To understand why new techniques have been adopted to meet this situation it is necessary to consider how frequency affects the currents flowing in conductors and dielectrics, and in particular it is necessary to examine the changes in ohmic and radiation resistances which occur when the frequency becomes very high.

1.1.1. *Skin Effect.* It may be shown that in the great majority of cases the resistance of a passive element increases as the frequency increases. This is due to the fact that a current in a conductor of finite cross-section tends to increase in density towards the surface of the conductor as the frequency increases, thus effectively diminishing the cross-sectional area. This is the so-called "skin effect" which has been known for many years.[1]

For a conductor the surface of which is plane, the current density diminishes to $1/e$ of its surface value at a depth d, where

$$d = \sqrt{\frac{10^9 \rho}{4\pi^2 \mu f}} \text{ cm.} \quad . \quad . \quad . \quad (1.1)$$

and ρ = the specific resistance of the conductor in ohm. cm.
μ = the permeability relative to free space in henries/cm.
f = frequency in cycles/sec.

This formula holds for any part of a conductor in which the radius of curvature of the cross-section is large compared with d. Table I exhibits the values of d for several values of f in the case of a copper conductor, and it will be seen that for very high frequencies, even if wire of 1 millimetre diameter is employed, d is still small enough to justify the approximation given by equation (1.1). The value of ρ taken is $1{\cdot}74 \, . \, 10^{-6}$ ohm. cm.

TABLE I

f (c/s)	10^6	10^7	10^8	10^9	10^{10}
d (mm.)	0·066	0·021	0·007	0·002	0·001

As will be seen in the sequel, a knowledge of the skin depth is useful in designing circuits for very short wave-lengths. Since the current is confined to a relatively thin layer at the surface of a conductor, it is possible to use for such circuits almost any material provided that it can be plated at the appropriate parts with a good conductor. As may be seen from Table I, a plating a thousandth of an inch thick should be effective for all frequencies above 100 Mc/s.

1.1.1.1. *The Resistance of an Isolated Cylindrical Conductor.* The increase due to skin effect in the resistance of a straight cylindrical conductor isolated from all other conductors may be calculated,[(2)] and the result is shown in Table II where the ratio of the high frequency resistance R to the D.C. resistance R_0 is given as a function of the parameter x, where

$$x = 2\pi r \sqrt{\frac{2\mu f}{10^9 \rho}}$$

Here μ, f, ρ have the same meanings as in equation (1.1) and r is the radius of the wire in centimetres.

TABLE II

x	1·0	1·5	2·0	2·5	3·0	3·5	4·0
R/R_0	1·0052	1 0258	1·0782	1·1754	1·3181	1·4920	1·6779
x	4·5	5·0	6·0	7·0	8·0	10·0	15·0
R/R_0	1·8628	2·0427	2·3936	2·7432	3·0945	3·7986	5·5622
x	20·0	25·0	30·0	40·0	60·0	100	
R/R_0	7·3277	9·0941	10·861	14·395	21·465	35·607	

To quote an example of the importance of this phenomenon, the resistance of a straight piece of No. 20 S.W.G. copper wire at 100 Mc/s is about 34 times its D.C. resistance, i.e. about 0·9 ohm/metre. Since the skin depth is small compared with the radius of the wire, the radius must be multiplied by a factor R/R_0 to bring the resistance back to its D.C. value, and therefore to obtain the equivalent at 100 Mc/s wire of diameter 34 × 0·0914 cm. or 3 cm. must be used.

It should be remarked that Table II is only true for sinusoidal oscillations and for wires which may be regarded as being at an infinite distance from other conductors.

When the skin effect is such that the resistance per unit length may be regarded as proportional to the radius of cross-section, an approximate formula for the high frequency resistance, due to Rayleigh[3] may be used. This is

$$R = \frac{1}{r}\sqrt{10^{-9}f\rho} \text{ ohms} \quad . \quad . \quad . \quad (1.2)$$

where r, f, ρ have the previous meanings. This is the formula generally used at ultra high frequencies for non-magnetic metals.

1.1.1.2. *The Resistance of a Pair of Parallel Wires.* If each wire has radius r and if their axes are distance d apart, then the formula for the high frequency resistance is[4]

$$R = \frac{2}{r}\sqrt{\frac{10^{-9}f\rho}{\left(1 - \frac{4r^2}{d^2}\right)}} \text{ ohms/cm.} \quad . \quad . \quad (1.3)$$

It will be observed that where $d \gg r$, this is simply twice the resistance of each wire, so the pair is regarded as a "go" and "return." The assymetrical distribution of current on each wire due to the proximity of the other leads to the correction factor $(1 - 4r^2/d^2)$. Applying this formula to the case of a metre of twin cable, each wire of which has a diameter of 1 mm., the distance between centres being 1 cm., the resistance at 100 Mc/s is 1·68 ohms.

1.1.1.3. *Coaxial Cylinders.* In this case there is no correction to the use of formula (1.2), and so, if r_1 = outer diameter of the inner conductor and r_2 = inner diameter of outer conductor

$$R = \sqrt{10^{-9}f\rho}\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \quad . \quad . \quad . \quad (1.4)$$

1.1.1.4. *Variation of Inductance with Frequency.* The fact that high frequency currents tend to flow near the surface of a conductor causes the inductance of the latter to decrease as the frequency increases. This produces a great simplification of the formulae for the inductances of lines in the region where equations (1.3) and (1.4) apply to the resistance. So long as the latter is not large, the inductance of a pair of parallel wires per cm. may be written

$$L_P = 9{\cdot}21 \,.\, 10^{-9} \log_{10}\frac{d}{r} \text{ henry/cm.} \quad . \quad . \quad . \quad (1.5)$$

and the corresponding equation for coaxial cylinders is

$$L_C = 4{\cdot}61 \,.\, 10^{-9} \log_{10}\frac{r_2}{r_1} \text{ henry/cm.} \quad . \quad . \quad . \quad (1.6)$$

1.1.2. *Radiation Resistance.* A varying electric or magnetic field by its very nature is in general propagated throughout the adjoining medium, but until the very high frequencies are reached, the radiation from small conducting elements is comparatively small. It is true that the receiver engineer finds it necessary at all frequencies to prevent stray radiation from reaching his circuits, but this is because he is concerned with comparatively minute powers. At frequencies below 100 Mc/s it is unnecessary to consider the power radiated from a circuit element in relation to the current flowing in it, unless in the case of a structure specifically designed to radiate. In fact at the lower frequencies it is somewhat of a problem to design an aerial which will effectively discharge radio power into space.

Far different is the case at the ultra high frequencies. There it is a continual preoccupation of the engineer to prevent power from leaking away from the conducting paths along which he wishes it to flow.

The radiation from a circuit element is measured in terms of its "radiation resistance," R, where R is given by the equation $W = RI^2$ and W is the power radiated when a current I flows in the element. It is impossible to give a general formula for R even in the simple case of a short element of wire, since the solution depends upon the nature of the surrounding conductors. But for an element of length l in which the current distribution is uniform and which is isolated from other conductors, the resistance is

$$R = 80\pi^2\left(\frac{l}{\lambda}\right)^2 \quad . \quad . \quad . \quad . \quad (1.7)$$

where λ is the wave-length of the oscillatory current. Changing the position of the element or its mode of excitation changes the value of the numerical constant in R, but the latter is always proportional to $(l/\lambda)^2$. This explains why it is impossible to have even the shortest lengths of unshielded conductor at ultra high frequencies. In the case of a pair of parallel wires, the radiation is approximately proportional to $(d/\lambda)^2$ where d has its usual meaning. This is not inappreciable for almost any practical Lecher wire system at centimetric wave-lengths, and so the use of open lines is limited to a comparatively narrow band of frequencies.

1.1.3. *Dielectric Losses.* Unfortunately for the engineer, very few materials which are otherwise suitable for use as dielectrics are free from losses at ultra high frequencies. Of the common dielectrics in use at 1–10 Mc/s, only a few survive tests at 100 Mc/s. Most of the phenol products drop out at comparatively low frequencies. The steatites remain useful up to several hundreds of megacycles. But at the centimetre-wave end of the spectrum only certain glasses, mica and certain ceramics are tolerable, and the only really good dielectrics are the polystyrenes. One of the most important avenues for research at the moment is in the exploration of dielectric materials, for, although the polystyrenes have good electrical properties, their low melting point is very troublesome. Polytetrafluorethylene (P.T.F.E.) is much better in this respect, but its present high cost prohibits its use in all except laboratory apparatus.

Some typical figures (measured at a frequency of 25,000 Mc/s) are given in Table III both for the dielectric constant and power factor (tan θ). For good dielectrics, the figures do not alter radically as the frequency is diminished. For example, polystyrene is only half as lossy at a frequency of 1 Mc/s.

TABLE III

Material	K	tan θ	Material	K	tan θ
Polystyrene	2·55	0·0008	Silica	3·85	0·0001
Perspex	2·65	0·012	Paraffin Wax	2·26	0·0001
Ebonite	2·73	0·004	Plastazote	1·16	0·003
707 Glass (Corning)	4·0	0·0024	Frequentite	6·0	0·003
L1 Glass	6·8	0·009	Tempradex	15·1	0·0002

It will be obvious that the ultra high frequency engineer has to make increasing use of air as a dielectric and cannot rely on finding insulants for mechanical support. This limitation on design has to some extent determined techniques. Whether or not it is a fundamental limitation time alone will tell.

1.1.6. *Inductance at U.H.F.* The physical realization of an inductance becomes a matter of some difficulty as the frequency increases, and that for two reasons. First, as has been explained, the ohmic and radiation resistances of conventional coils become prohibitively large, and second, the self-capacity of such a coil may easily swamp the inductive component. In fact, although the concept of inductance is still useful at the highest frequencies, L is no longer directly measurable. As will be shown in the sequel (1.2), an inductive impedance can be achieved by the use of a line with distributed parameters. Of course, if we were prepared to make and employ exceedingly small components, coils could be used at much higher frequencies. But the necessary restriction in this direction, if it be not size as such, is the power handling capacity of the component. Scaling down in length means scaling down in other things as well, and that is not acceptable to the electrical engineer.

1.1.7. *Capacitance at U.H.F.* To a lesser degree what has been said of inductance applies to capacitance, but in this case the associated series inductance is one of the limiting factors. The distribution of current throughout the "condenser" is also important and gives to certain conventional circuit elements a different significance at U.H.F. As in the case of inductance, a line may be used (see 1.2) to obtain a capacitative impedance at any frequency, and it is to the line that the engineer must look for his practical condenser. However, owing to the nature of a capacity, conventional condensers can be used at higher frequencies than conventional coils. In fact, if simple forms of parallel plate condensers are used, they may be successfully identified with capacities at frequencies up to 1000 Mc/s.

1.1.8. *Resistance at U.H.F.* A most important element in radio design is the non-reactive resistance, for, although it is not usually required in transmitters and receivers as such, all measurements are based upon a dissipative element which is not frequency sensitive. At U.H.F. it is almost obvious that such an element cannot take the form of a wire or carbon resistance of the conventional type. In the investigation of metric and centimetric waves the line with distributed parameters again offers the solution, for a lossy concentric line forms an ideal U.H.F. ohmic resistance under certain circumstances.

The alternative resistance which finds considerable favour at U.H.F. is the water load. If the oscillatory energy is properly presented to a mass of water, the latter can be made to absorb the energy in a manner which allows of accurate measurement. The water calorimeter is one of the few absolute instruments which can be employed at U.H.F. to measure power.

1.2. **Parallel Wires and Concentric Lines.**[4] It is the purpose of this paragraph to show that properly terminated lines with distributed parameters can be made to simulate any arbitrary arrangement of "lumped" components, and that such lines can replace "lumped" circuits at U.H.F. with considerable diminution of the ohmic and radiation resistances which have been shown to be inevitable with conventional coils and condensers.

1.2.1. *Theory.* The particular case of the concentric line is the more important, but the analysis given below may be applied to any line in which the radiation resistance can be neglected.

Let the distance measured along the line be x and let L, C, R and G represent the inductance, capacity, resistance, and leakage conductance respectively of unit length of the line. Then if δV is the change of voltage along the line in δx associated with a change of current δI, by the elementary laws of electromagnetism

$$\left.\begin{aligned} \delta V &= \left(-L\frac{\partial I}{\partial t} - RI\right)\delta x \\ \text{and} \qquad \delta I &= \left(-C\frac{\partial V}{\partial t} - GV\right)\delta x \end{aligned}\right\} \quad . \quad . \quad . \quad (1.8)$$

Whence $$\frac{\partial V}{\partial x} = -(R + j\omega L)I$$

and $$\frac{\partial I}{\partial x} = -(G + j\omega C)V$$

assuming that both V and I have the form $f(x)e^{j\omega t}$ where $f(x)$ is an arbitrary function of x.

Denote $R + j\omega L$ by Z = series impedance per unit length

and $G + j\omega C$ by Y = shunt admittance per unit length

Therefore $$\left.\begin{aligned}\frac{\partial V}{\partial x} &= -IZ \\ \frac{\partial I}{\partial x} &= -VY\end{aligned}\right\} \qquad (1.9)$$

and

whence $$\frac{\partial^2 V}{\partial x^2} = YZV \qquad (1.10)$$

The solution is $$V = \{Ae^{\sqrt{YZ}\,x} + Be^{-\sqrt{YZ}\,x}\}e^{j\omega t} \qquad (1.11)$$
$$= V_x e^{j\omega t}$$

To evaluate A and B, let $V_x = V_s$ and $I_x = I_s$ at $x = 0$. Then, substituting in (1.9) and noting that

$$I = -\frac{1}{Z}\frac{dV}{dx}$$

we have $$\left.\begin{aligned}A &= \tfrac{1}{2}\left\{V_s - \sqrt{\frac{Z}{Y}}I_s\right\} \\ B &= \tfrac{1}{2}\left\{V_s + \sqrt{\frac{Z}{Y}}I_s\right\}\end{aligned}\right\} \qquad (1.12)$$

and

Write $\sqrt{Z/Y} = Z_0$ and $\sqrt{YZ} = \gamma$

Therefore $$\left.\begin{aligned}V &= V_s \cosh \gamma x - Z_0 I_s \sinh \gamma x \\ I &= I_s \cosh \gamma x - (V_s/Z_0) \sinh \gamma x\end{aligned}\right\} \qquad (1.13)$$

and

omitting the term $e^{j\omega t}$

These are the general equations giving the solutions for V and I in the case of a line for which the parameters are known. Special cases of this will now be considered.

(i) Let $x \to \infty$. Then as $x \to \infty$, $V \to I \to 0$, since the line is lossy.

But $V = \frac{V_s - Z_0 I_s}{2} e^{\gamma x} + \frac{V_s + Z_0 I_s}{2} e^{-\gamma x}$

from (1.11) and (1.12).

Hence $V_s - Z_0 I_s = 0$ or $V_s/I_s = Z_0$.

The input impedance in the case of an infinite line is known as the "characteristic impedance" of the line (Z_0).

Also, since $V_s - Z_0 I_s = 0$,

$$V = V_s e^{-\gamma x} \text{ and } I = I_s e^{-\gamma x} \qquad . \quad . \quad . \ (1.14)$$

These equations give γ the significance of a "propagation constant," the value of which in terms of the line parameters is

$$\gamma = (\alpha + j\beta) = \sqrt{(R + j\omega L)(G + j\omega C)} \qquad . \ (1.15)$$

(ii) Let x be finite and equal to l. Then re-writing (1.13)

$$\left.\begin{aligned} V_l &= V_s \cosh \gamma l - Z_0 I_s \sinh \gamma l \\ I_l &= I_s \cosh \gamma l - (V_s/Z_0) \sinh \gamma l \end{aligned}\right\} \qquad . \quad . \quad . \ (1.16)$$

Consider the case where the ohmic losses in the length l are negligible compared with L and C.

Then $$\gamma = j\omega\sqrt{LC} \qquad . \quad . \quad . \quad . \ (1.17)$$

If now the line is short-circuited at $x = l$, then $V_l = 0$

and $$V_s \cosh \gamma l = Z_0 I_s \sinh \gamma l$$

or $$\frac{V_s}{I_s} = Z_{input} = Z_0 \frac{\sinh \gamma l}{\cosh \gamma l}$$

$$= jZ_0 \tan (\omega l \sqrt{LC}) \qquad . \quad . \ (1.18)$$

Hence the input impedance is variable from $+\infty$ to $-\infty$ as in the ideal case of the non-lossy lumped circuit consisting of a condenser in parallel with a coil. By choosing the value of l we can regard the length of line as an inductance or a capacity of any value.

(iii) Consider now the product LC which determines γ in the case of a non-lossy line.

It may be shown[5] that the inductance of a concentric line per unit length is given by

$$L = 2 \log_e \frac{r_2}{r_1} \text{ e.m.u./cm.} \qquad . \quad . \quad . \ (1\ 19)$$

and the capacitance per unit length by

$$C = \frac{1}{2c^2 \log_e (r_2/r_1)} \text{ e.m.u./cm.} \qquad . \quad . \quad . \ (1.20)$$

where c is the velocity of light in cm./sec.

Therefore $LC = 1/c^2$ and $\gamma = j\omega\sqrt{LC} = j\omega/c = 2\pi j/\lambda$ where λ is the free space wave-length corresponding to a frequency f and a pulsatance ω,

i.e.
$$c/\lambda = f = \omega/2\pi$$

In the more general case, where the dielectric between the conductors has a dielectric constant k, the formula for the capacitance per unit length becomes

$$C_k = \frac{k}{2c^2 \log_e r_2/r_1}$$

and
$$\gamma = j\omega\sqrt{k}/c = 2\pi j\sqrt{k}/\lambda \quad . \quad . \quad (1.21)$$

(iv) Applying this value of γ (with $k = 1$) to the non-lossy line, equation (1.18) for the input impedance becomes

$$Z_{input} = jZ_0 \tan 2\pi\left(\frac{l}{\lambda}\right)$$

and Z_{input} is zero, when $2l = n\lambda$, for $n = 0, 1, 2, \ldots$ Similarly, Z_{input} is infinite when $2l = (n + \frac{1}{2})\lambda$, for $n = 0, 1, 2, \ldots$

In the same way, if the line is open-circuited at $x = l$ then $I_l = 0$ and

$$Z_{input} = Z_0 \coth \gamma l$$

and if the line is non-lossy, $Z_{input} = jZ_0 \cot 2\pi\left(\frac{l}{\lambda}\right)$

So Z_{input} is zero when $2l = (n + \frac{1}{2})\lambda$.

And Z_{input} is infinite when $2l = n\lambda$.

(v) In the general case where γ contains a real term, from (1.16) it may be deduced that

$$Z_s = Z_0 \frac{Z_l \cosh \gamma l + Z_0 \sinh \gamma l}{Z_0 \cosh \gamma l + Z_l \sinh \gamma l} \quad . \quad . \quad (1.22)$$

where Z_s is the input impedance.

If $Z_l = 0$, i.e. if the line is short-circuited at l,

$$Z_s = Z_0 \tanh \gamma l$$

Write $\tanh \gamma l = \dfrac{e^{(\alpha + j\beta)l} - e^{-(\alpha + j\beta)l}}{e^{(\alpha + j\beta)l} + e^{-(\alpha + j\beta)l}} = M(\cos\theta + j\sin\theta)$

Then, writing each term in the proper form, we have

$$M = \frac{\sqrt{\sinh^2 \alpha l \cosh^2 \alpha l + \sin^2 \beta l \cos^2 \beta l}}{\cos^2 \beta l \cosh^2 \alpha l + \sin^2 \beta l \sinh^2 \alpha l}$$

and
$$\theta = \tan^{-1} \frac{\sin \beta l \cos \beta l}{\sinh \alpha l \cosh \alpha l}$$

Simplifying,

$$M = \sqrt{\frac{\cosh 2\alpha l - \cos 2\beta l}{\cosh 2\alpha l + \cos 2\beta l}} \quad . \quad . \quad . \quad . \quad (1.23)$$

and

$$\theta = \tan^{-1}\left(\frac{\sin 2\beta l}{\sinh 2\alpha l}\right) \quad . \quad . \quad . \quad . \quad . \quad (1.24)$$

(vi) The values of α and β will now be found in terms of the line parameters.

$$\gamma = \alpha + j\beta = \sqrt{YZ} = \sqrt{(R + j\omega L)(G + j\omega C)}$$

In many practical cases and in almost all lines used at U.H.F., the leakage conductance G can be neglected.

Therefore $\alpha + j\beta = j\omega\sqrt{LC}\left\{1 - j\frac{R}{\omega L}\right\}^{\frac{1}{2}}$

Let $\omega L/R$ be represented by Q. As in the case of lumped circuits, $1/Q$ is a measure of the losses.

Then $$\alpha^2 - \beta^2 + 2j\alpha\beta = -\omega^2 LC(1 - j/Q)$$

giving $$\alpha^2 - \beta^2 = -\omega^2 LC$$

and $$2\alpha\beta = \omega^2 LC/Q$$

Clearly, to the degree of approximation usually employed,

$$\alpha = \frac{\omega\sqrt{LC}}{2Q} \text{ and } \beta = \omega\sqrt{LC}$$

But it has already been shown that $LC = 1/c^2$,

whence $$\alpha = \frac{\omega}{2Qc} = \frac{2\pi}{2Q\lambda} = \frac{\pi}{Q\lambda} \quad . \quad . \quad . \quad (1.25)$$

and $$\beta = \frac{2\pi}{\lambda} \quad . \quad . \quad . \quad (1.26)$$

(vii) Returning now to equation (1.23), which gives the ratio of the moduli of the input to the characteristic impedance, i.e. $|Z|/|Z_0|$ in the case of a line short-circuited at $x = l$, it is interesting to calculate M for a sample line. Consider the case where the frequency of the oscillation is 300 Mc/s ($\lambda = 100$ cm.) and the Q of the line is 10.

Then $\alpha = \pi/1000$ and $\beta = 2\pi/100$ from equations (1.25) and (1.26). Figure 1 exhibits the variation of M with l. It will be seen that the input impedance oscillates about the value of the characteristic impedance with decaying amplitude. The curve asymptotes to $M = 1$. Reference will be made to this figure from time to time, for it brings to light several important qualities of the short-circuited line. A very similar curve may be drawn for the open-circuited line.

(viii) Consider again the important parameter $Q = \omega L/R$.

From equations (1.19) and (1.4)

$$Q = 4\pi r_2 \sqrt{f/10^9 \rho}\, \frac{\log_e r_2/r_1}{1 + r_2/r_1} \quad . \quad . \quad . \quad (1.27)$$

As the ratio r_2/r_1 increases from slightly greater than unity to infinity the quantity

$$\frac{\log_e r_2/r_1}{1 + r_2/r_1}$$

increases to a maximum and then diminishes again.

Hence, *for a given value of* r_2, Q has a maximum when

$$\log_e r_2/r_1 = 1 + \frac{r_1}{r_2}$$

i.e. when $r_2/r_1 = 3{\cdot}591$

If this value is inserted in equation (1.27),

$$Q = 4\pi r_2 \sqrt{f/10^9 \rho}\, \frac{\log_e 3{\cdot}591}{1 + 3{\cdot}591}$$

or, for copper

$$Q = 0{\cdot}0838 r_2 \sqrt{f} \quad . \quad . \quad . \quad . \quad (1.28)$$

where r_2 is in cm. and f in cycles/sec.

(ix) Finally, consider the input impedance of a quarter wave line. In this case $l = \lambda/4$ and in equation (1.23),

$$M = \coth \alpha l$$

But $\alpha = \pi/Q\lambda$ and $l = \lambda/4$,

Therefore $M = \coth(\pi/4Q) = \dfrac{1}{\tanh(\pi/4Q)}$

Now $\tanh x \doteq x$ for small values of the latter

Therefore $M \doteqdot 4Q/\pi$

and $Z_{input} = 4QZ_0/\pi \quad . \quad . \quad . \quad . \quad (1.29)$

This is a very useful formula in many cases. Where Q is large enough it can be extended to lines of length $3\lambda/4$, $5\lambda/4$, etc.

Summing up the results of this paragraph, it has been shown that a line with distributed parameters can be made to behave electrically in exactly the same way as an arbitrary "lumped circuit," or, putting it another way, every property possessed by an arbitrary network of passive lumped circuit elements can be exactly reproduced by a branched line. In addition, a concentric line has many other properties of which use can be made, the most important of which is the recurring nature of the phenomena along its length. Although it is difficult in practice to make a concentric line having a Q equal to that given by theory (equation 1.27), nevertheless the practical Q's are many times larger than can be obtained with lumped circuits. Using silver-plated cylinders of diameters 1 and 3·6 cm., an unloaded Q of 2000 is easily obtained in a quarter-wave length line at 400 Mc/s.

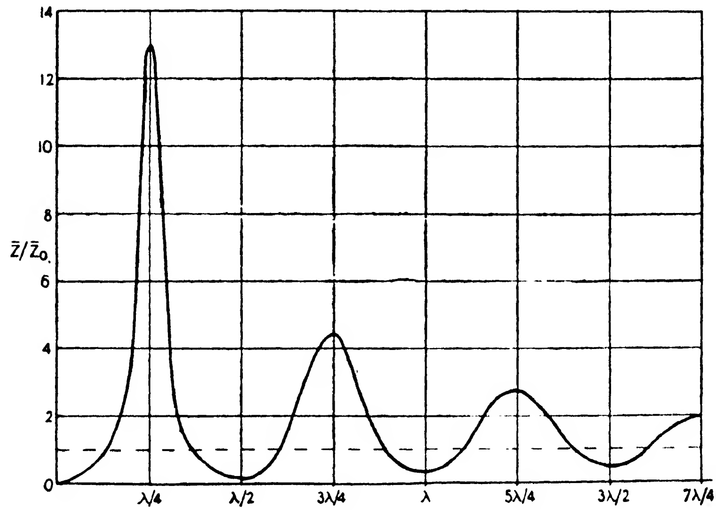


Fig. 1. The variation with length of the ratio of the modulus of the input impedance to the characteristic impedance in the case of a line for which $Q = 10$

This forms an excellent basis for transmitter and receiver circuits, where losses must be kept to a minimum, corresponding, as it does, to an input impedance of about 2500 Z_0 where Z_0 is the characteristic impedance. The reader should study the case of the open-circuited line in the manner adopted above for the short-circuited line. Corresponding formulae will, of course, be obtained. It will also be advantageous to go over the reasoning of this paragraph, having in mind a parallel-wire line rather than a concentric line. It will be found that all the formulae are unaltered with the exception of those which involve L, C, R or $\sqrt{L/C}$. Formulae involving $\sqrt{LC}$ are unaltered. Figure 1 applies equally to a parallel-wire line, but it must be remembered that in evaluating Q in this case account must be taken of the radiation losses as well as the ohmic losses.

1.2.2. *Practice.* When lumped circuits have become impracticable, the tendency is to use parallel-wire lines and only to resort to concentric lines where the radiation resistance of the parallel wires becomes important. Lumped circuits can be used in receivers at higher frequencies than in transmitters, for the power-carrying capacity of the circuits need not be so large. An extreme case of this is to be found in certain German search receivers used during Word War II. There, by employing condensers which formed an integral part of the receiver chassis and coils of very small dimensions, lumped circuits were used at frequencies up to 1000 Mc/s. In general, however, if high selectivity is required, coil and condenser techniques must be considerably modified at frequencies above 200 Mc/s.

1.2.2.1. *Amplifier and Oscillator Circuits.* In transmitters lines are usually more efficient than coils at frequencies above 100 Mc/s. The typical case of a small power amplifier designed to cover from 100 to 200 Mc/s will be considered. A suitable valve would be a double beam tetrode with an anode to anode capacitance of the order of $4\mu\mu$F. In this case a line about 20 cm. long is required. An illustration of a practical case of this is shown in Figure 1, Plate I. Here the open line has been bent into an arc of a circle. This economizes in space and allows a simple circular tuning mechanism to be employed. The latter consists of a shorting bar, mounted on springs at the end of a radial arm. This slides round the circle formed by the parallel conductors. In the illustration it will be noted that the power is abstracted from the line circuit by means of condensers, mounted near the valve end of the circuit. Double-ended or balanced output is obtained by two such condensers, one on each conductor. Unbalanced or single-ended output is obtained by using one condenser to abstract the power and the other to compensate partially for the unbalance produced. In practice such an arrangement works very well over a surprisingly large frequency band.

In designing such a parallel-wire circuit, the procedure is as follows—

(i) Determine the minimum length of line (l_0) which it is necessary to have outside the valve envelope at the highest frequency of the range (f_0).

(ii) Find, experimentally, the characteristic impedance Z_0 which will just fulfil condition (i).*

(iii) This determines the ratio d/r where d is the distance between centres of conductors and r is the equivalent circular radius of the conductors. Now employ the formula for the Q of a parallel wire system

$$Q = \frac{\omega L}{R} = \frac{2\pi f \,.\, 9{\cdot}21 \,.\, 10^{-9} \log_{10}(d/r)}{2\sqrt{10^{-9} f\rho/r} + 80\pi^2 f^2 d^2/c^2}$$

which includes the ohmic and radiation resistances.

* For the theoretical method, see Chapter II, equation 2.1.

C

This equation is of the form

$$Q = \frac{A \log_{10} M}{B/r + CM^2r^2}$$

and Q has a maximum value when $r^3 = B/2CM^2$, or $r = B/2Cd^2$.

This determines r and hence d, and the line is defined. The distortion of the electrostatic field due to the proximity of the conductors may usually be neglected in this calculation, and if conductors of square cross-section are employed, as is generally the case, $2\pi r$ may be written equal to $4a$ where a is the length of one side of the square. This is justified by the "skin effect," for the current carrying capacity of the conductor is proportional to the perimeter of its cross-section.

The most important practical difficulty in using line circuits lies in ensuring good electrical and mechanical contact between the line elements and the tuning element. The attempt to overcome this has given rise to many ingenious devices and to the use of many different metals at the area of contact. The problem has been satisfactorily solved for transmitters but not for receivers. In the former, electrical noise due to moving contacts does not occur, since the machine is not in operation while it is being tuned. This is not always true of receivers, and so far no moving contact has been invented which is not associated with this electrical noise. The trouble may be eliminated by arranging the moving contact to be at a part of the circuit where high frequency currents do not flow, but this solution is almost impossible to achieve where wide-band tuning is required. A simple example of a sliding contact is shown in Figure 1, Plate I. Here the material of the shorting bar is bronze-graphite and the line conductors are of brass heavily silver-plated. Figure 2, Plate I, shows the various elements of a concentric line suitable for use over the frequency range 200 to 400 Mc/s in conjunction with a disc-seal triode valve (see Chapter II). The contacts in this case are spring fingers of beryllium copper, silver plated, and sliding against silver-plated cylinders. So long as a smooth finish is obtained on the rubbing surfaces, such contacts are entirely satisfactory over long periods of time. It must again be emphasized that they are *not* satisfactory in receivers which must operate while tuning is in progress.

1.2.2.2. *Wave-meters.* One of the most interesting applications of the concentric line is to wave-meters. As shown by the theory of paragraph 1.2.1, an open-ended quarter-wave line has a very high resistive shunt impedance, or, in more usual parlance, the open-ended quarter-wave line is a resonator at that wave-length. The line also resonates at $l = 3\lambda/4$, $5\lambda/4$, etc. Therefore, if a concentric line is made which is short-circuited at one end and open-circuited at the other, and in which the length of the inner conductor can be varied and measured, it is possible to measure the distance between two successive impedance maxima. In this way the instrument can be used as a wave-meter. In practice it is necessary, of course, to arrange means whereby a small amount of power can be fed into the line and to obtain relative measurements of the input impedance by loosely coupling a detector. A typical arrangement is shown diagrammatically in Figure 2, all measuring equipment being omitted.

A is the inner conductor, free at one end, and making contact with the outer conductor B at the short-circuited end by means of

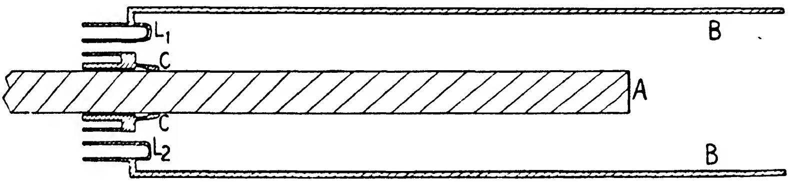


Fig. 2. Axial section of a wave-meter consisting of an open-ended concentric line

a sliding contact C. At this end two small loops of wire L_1 and L_2 are inserted into the cavity to supply the energy from the source and to measure the effect of the line in a detector.

As the length of A inside the conductor B is increased from zero, the detector output varies in the manner shown in Figure 1, having maxima at points $\lambda/2$ cm. apart. But Figure 1 was drawn for a Q of 10, whereas in a good wave-meter a working Q of 1000 is not difficult to attain. This creates an enormous sharpening of the resonance maxima. Figure 3 is drawn for the range $l = 0{\cdot}23\lambda$ to $l = 0{\cdot}27\lambda$ for a Q of 1000. The value of $|Z|/Z_0$ for the first maximum calculated by equation 1.29 is 1274. The corresponding values for the second and third maxima are 425 and 255.

Many ingenious modifications of the fundamental structure suggested by Figure 2 have been made from time to time. Owing to the electrical excellence of such meters, the mechanical problem of measuring the position of the inner conductor becomes considerable. Methods of eliminating back-lash and of shortening the travel of the inner conductor are important. Figure 3, Plate I, shows a wave-meter designed by the author for the measurement of wavelengths between 3 and 30 cm. Here advantage is taken of the force of gravity to avoid back-lash, and the meter scale is so arranged that the instrument works on the $\lambda/4$ resonance between 10 and 30 cm. and on the $3\lambda/4$ resonance between 3 and 10 cm. This can be done without ambiguity, if the necessary checks are made on other resonance points. The advantage obtained is that the whole scale is only 5 cm. long. Readings can be taken to one-hundredth of a millimetre if the accuracy of the screw motion justifies it.

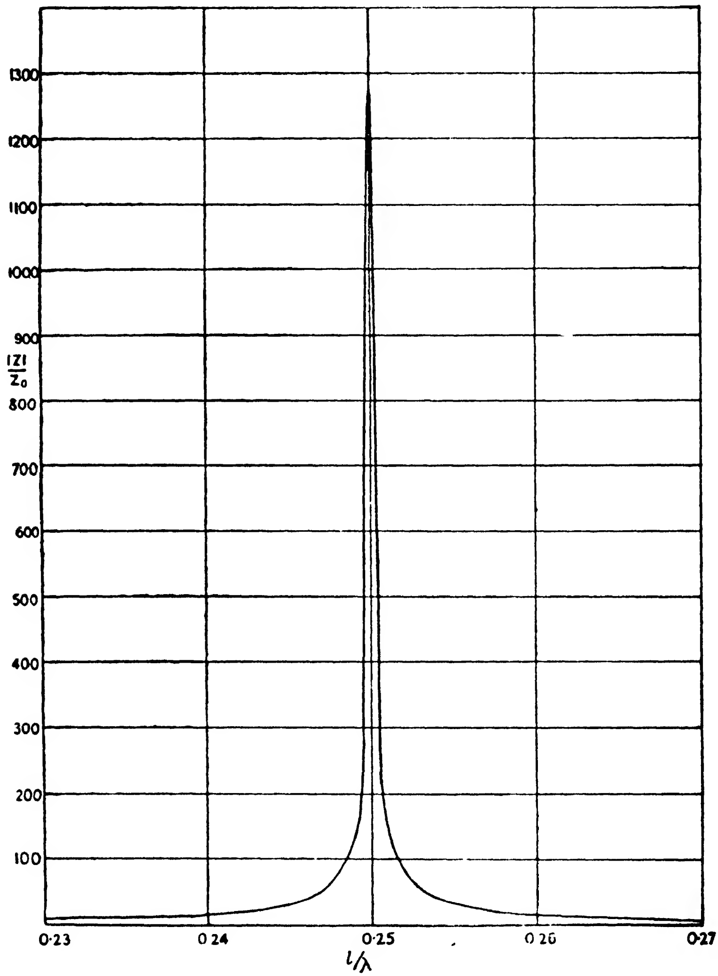


Fig. 3. The variation with length of the ratio of the modulus of the input impedance to the characteristic impedance for a line of $Q = 1000$

1.2.2.3. *Matching Stubs.* It has been shown that a section of line can be made to give almost any shunt impedance, but in the two preceding practical cases only the resonant shunt impedance has been used.

In many U.H.F. circuits, where power is to be transferred from one point to another, it is necessary to arrange that the impedance at a given point in a line has a given value. One of the more obvious cases is where a power measurement has to be made, using a dissipative element of unknown or variable resistance, or where the resistance is for other reasons fixed at an inconvenient value. For example, where power measurement is made using a wire as the resistive element, considerations of total power dissipation frequently determine the resistance of the element. In this case also, the resistance varies with temperature, a fact which may or may not be made

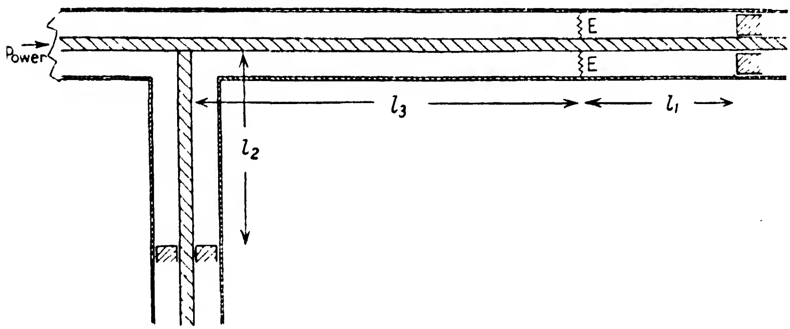


Fig. 4. Axial section of a matching unit comprising a series and a parallel stub line

use of in making the measurement, but which certainly complicates the matching problem. This problem is exhibited in Figure 4.

If the power to be measured flows in from the left towards the dissipative element E of resistance R, it is always possible to arrange two stub lines to match R to the characteristic impedance of the line, Z_0. The first stub is connected directly across R. Let its length to the short-circuited end be l_1. The second, also short-circuited, and of length l_2, is connected across the line a distance l_3 away from E. Two stubs are necessary, giving two degrees of freedom, because, although a section of line can have any *complex impedance*, there is a necessary relation between the resistive and reactive components, and so, to achieve any arbitrary value of each, two sections of line must be used. In the above arrangement l_1, l_2 are considered variable and l_3 fixed. By carefully choosing l_3, it is possible to make l_1 and l_2 always within the range $O < \frac{l_1}{l_2} < \lambda/4$ cover all possible variations of R. Another point is worthy of notice. Where the coupling between the line and the generator can be varied, it is possible to match E into the line using only one stub. In this case the variation of the coupling gives the second degree of freedom.

1.3. **Wave-Guides and Resonators.**[6] Ohmic losses in concentric lines used as circuits begin to be troublesome at centimetric wave-lengths. For example, a copper conductor of diameter 1 cm. has a resistance per metre of ½ ohm at 3000 Mc/s. Dielectric losses, as has been indicated, also become relatively high at these wave-lengths, and as it is impossible in many cases to support the inner conductor of a concentric line without solid dielectric, for this reason also it is desirable to find some other form of circuit. It is fortunate, therefore, that a different form of guide for electro-magnetic energy begins to become a practical proposition at just about this region of the spectrum. A metal tube, the internal diameter of which is of the same order as the wave-length, possesses many of the properties of a concentric line. To be more precise, such a tube will transfer power from one end to the other with little loss—it will guide the power along its length; when properly ter-minated it will behave like an inductive or capacitative impedance; and when closed or open-circuited at one or both ends, it will resonate when one of its dimensions bears a certain relationship to the wave-length. In fact, where such tubes or guides differ from concentric lines in their properties, it is due only to the fact that the linear dimensions of their cross-section are of the same order of magnitude as the wave-length. This, for example, gives rise to the various modes of oscillation in wave-guides, all of which except one are suppressed in practical concentric lines where the radii of the inner and outer conductors are small compared with the wave-length.

It is proposed, first, to consider some simple aspects of the theory of wave-guide propagation and of cavity resonators, and second, to describe certain practical circuits which employ guides and cavities at U.H.F.

1.3.1. *Theory.* The theoretical approach to lines was made *via* the conceptions of inductance, capacitance and resistance per unit length, the equations determining the properties of the lines being those associated with conventional high frequency engineering. If, instead of using this particular mathematical technique, the more general equations for the electromagnetic field are employed, a wider theory results, and one which includes all forms of hollow guide as well as parallel wire and concentric guides. In the present section it is proposed to develop the elementary theory of tubular guides, using the scalar form of the electromagnetic equations. This approach, while not so elegant as the method of vectors,[7] is perhaps less likely to lead the reader into unknown mathematical fields.

The classical Maxwell equations, linking the components of electric field strength E_x, E_y, E_z with the components of magnetic field strength H_x, H_y, H_z may be written for a non-conducting medium in the form*

$$\left.\begin{aligned}
\frac{K}{c}\frac{\partial E_x}{\partial t} &= \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} \quad . \text{ (A)} \\
\frac{K}{c}\frac{\partial E_y}{\partial t} &= \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} \quad . \text{ (B)} \\
\frac{K}{c}\frac{\partial E_z}{\partial t} &= \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \quad . \text{ (C)} \\
-\frac{\mu}{c}\frac{\partial H_x}{\partial t} &= \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \quad . \text{ (D)} \\
-\frac{\mu}{c}\frac{\partial H_y}{\partial t} &= \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \quad . \text{ (E)} \\
-\frac{\mu}{c}\frac{\partial H_z}{\partial t} &= \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \quad . \text{ (F)}
\end{aligned}\right\} \quad . \quad . \quad (1.29)$$

where μ and K are the permeability and dielectric constant of the medium respectively, and c is the velocity of light.[8] Let these equations be applied directly to the flow of electromagnetic energy along a rectangular tube, shown diagrammatically in Figure 5.

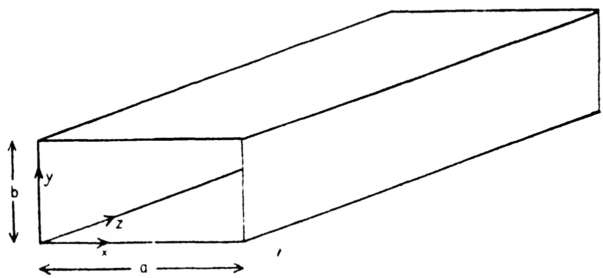


Fig. 5. A piece of rectangular wave-guide showing co-ordinate axes

Let the z-axis be along one edge of the tube, parallel to the axis, and let the x and y axes be the other edges. Then the analysis is concerned with the flow of electromagnetic energy along the tube, i.e. along the z-axis. But by differentiating equation 1.29 (C) with respect to t and substituting from D and E,

$$\frac{\mu K}{c^2}\frac{\partial^2 E_z}{\partial t^2} = \frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + \frac{\partial^2 E_z}{\partial z^2} - \frac{\partial}{\partial z}\left[\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right]$$

If the medium is non-conducting, then

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0$$

* Mixed electrostatic and electromagnetic units are used.

giving the wave-equation

$$\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + \frac{\partial^2 E_z}{\partial z^2} - \frac{\mu K}{c^2}\frac{\partial^2 E_z}{\partial t^2} = 0 \quad . \quad . \quad (1.30)$$

Similar equations can be developed for E_y, E_x, H_x, H_y, H_z. In the case under consideration, E_z may be written

$$E_z = XYZe^{j\omega t}$$

where X, Y, Z are functions of x, y, z. Equation 1.30 therefore becomes

$$\frac{1}{X}\frac{\partial^2 X}{\partial x^2} + \frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} + \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} + \frac{\mu K\omega^2}{c^2} = 0$$

If propagation occurs along the z-axis, $Z = e^{-\gamma z}$, and therefore

$$\frac{1}{X}\frac{\partial^2 X}{\partial x^2} + \frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} + k^2 = 0 \quad . \quad . \quad . \quad (1.31)$$

where
$$k^2 = \gamma^2 + \frac{\mu K\omega^2}{c^2} \quad . \quad . \quad . \quad . \quad (1.32)$$

The solution of (1.31) is

$$X = A\sin k_1 x + B\cos k_1 x$$
$$Y = C\sin k_2 y + D\cos k_2 y$$

Boundary conditions must now be introduced and this requires a consideration of the conducting medium forming the tube. It will be assumed for the purpose of this elementary discussion that the conductivity is infinite. Then there can be no finite electric force in the conductor, and $E_z = 0$ when $x = 0$ or $y = 0$.

Hence the cosine terms in X and Y disappear, and omitting the term $e^{j\omega t}$,

$$E_z = \sin k_1 x \sin k_2 y e^{-\gamma z}$$

where
$$k^2 = k_1^2 + k_2^2$$

E_z is also zero along the planes $x = a$ and $y = b$.

Whence $\sin k_1 a = 0$ or $k_1 = m\pi/a$, $m = 0, 1, 2, \ldots$

and $\sin k_2 b = 0$ or $k_2 = n\pi/b$, $n = 0, 1, 2, \ldots$

Hence
$$E_z = \sin(m\pi x/a)\sin(n\pi y/b)e^{-\gamma z} \quad . \quad (1.33)$$

where
$$\gamma^2 = (m\pi/a)^2 + (n\pi/b)^2 - \mu K\omega^2/c^2 \quad . \quad (1.34)$$

There are two possibilities emerging from (1.34). If γ^2 is positive, γ is real, and the wave is attenuated along the tube. This attenuation, it should be noted, is not due to losses. It is a diffraction phenomenon. If γ^2 is negative, the wave is not attenuated, but E_z has a periodic character which may be associated with a wavelength λ_g in the tube.

In this case

$$j^2\gamma^2 = \beta^2 = \mu K\omega^2/c^2 - (m\pi/a)^2 - (n\pi/b)^2$$

and

$$\lambda_g = 2\pi/\beta$$

From electromagnetic theory, the velocity of a wave[1] in a dielectric of constant K and permeability μ is v where

$$\frac{v}{c} = \frac{1}{\sqrt{\mu K}}, \text{ and } \frac{\mu K\omega^2}{c^2} = 4\pi^2 f^2/v^2 = \frac{4\pi^2}{\lambda^2}$$

Hence

$$\frac{1}{\lambda_g} = \left[\frac{1}{\lambda^2} - \left(\frac{m}{2a}\right)^2 - \left(\frac{n}{2b}\right)^2\right]^{\frac{1}{2}} \quad . \quad . \quad . \ (1.35)$$

The rectangular tube, therefore, of cross-sectional dimensions a and b acts as a guide for radiation the wave-length of which is less than λ_c, where

$$\lambda_c = \frac{2}{\{(m/a)^2 + (n/b)^2\}^{\frac{1}{2}}} \quad . \quad . \quad . \ (1.36)$$

The wave-length in the guide, when λ is less than λ_c, is given by equation (1.35).

Similar solutions can be found for E_x, E_y from the corresponding equations, but the simple cases, where $E_z = 0$ or $H_z = 0$ are most often found in practice.

1.3.1.1. *Transverse Magnetic* (*TM*) *Waves.* Let $H_z = 0$. Then the magnetic vector is in the *xy* or transverse plane, and Maxwell's equations become

$$E_x = \frac{c\gamma}{Kj\omega} H_y, \quad E_y = -\frac{c\gamma}{Kj\omega} H_x, \quad E_z = \frac{c}{Kj\omega}\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}\right)$$

and

$$H_x = \frac{-c}{\mu j\omega}\left(\gamma E_y + \frac{\partial E_z}{\partial y}\right), \quad H_y = \frac{c}{\mu j\omega}\left(\gamma E_x + \frac{\partial E_z}{\partial x}\right)$$

$$\frac{\partial E_y}{\partial x} = \frac{\partial E_x}{\partial y}$$

Substituting for H_x, H_y, the following expressions are obtained, where only functions of E_z appear on the right-hand side

$$E_x = -\frac{\gamma}{k^2}\frac{\partial E_z}{\partial x}, \quad E_y = -\frac{\gamma}{k^2}\frac{\partial E_z}{\partial y}, \quad H_x = \frac{Kj\omega}{ck^2}\frac{\partial E_z}{\partial y}, \quad H_y = -\frac{Kj\omega}{ck^2}\frac{\partial E_z}{\partial x}$$

But

$$E_z = \sin(m\pi x/a)\sin(n\pi y/b)e^{-\gamma z}$$

[1] To those who are not familiar with "Times New Roman," the type used throughout this book, it may be pointed out that the symbol here is the lower case italic "vee," the form by which velocity is always represented.

whence

$$\left.\begin{aligned}
E_x &= -j\frac{2\pi^2 m}{a\lambda_g k^2}\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{-2j\pi z/\lambda_g}\\
E_y &= -j\frac{2\pi^2 n}{b\lambda_g k^2}\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-2j\pi z/\lambda_g}\\
E_z &= \sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{-2j\pi z/\lambda_g}\\
H_x &= j\frac{2\pi^2 nK}{b\lambda k^2}\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-2j\pi z/\lambda_g}\\
H_y &= -j\frac{2\pi^2 mK}{a\lambda k^2}\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{-2j\pi z/\lambda_g}\\
H_z &= 0
\end{aligned}\right\} \quad . \quad (1.37)$$

The general transverse magnetic wave is represented by $TM_{m,n}$. The waves $TM_{0,n}$ and $TM_{m,0}$ do not exist, since for these values of m and n all components of the field vanish. The most important practical case is $TM_{1,1}$. The critical wave-length in this case is given by

$$\lambda_c = \frac{2}{\left\{\frac{1}{a^2}+\frac{1}{b^2}\right\}^{\frac{1}{2}}} = \frac{2ab}{\sqrt{a^2+b^2}}$$

For example, if $a = 4$ cm., $b = 3$ cm., $\lambda_c = 4{\cdot}8$ cm.

This is the "dominant" transverse magnetic wave. The actual values of the field components are found, as usual, by taking the real parts of the values given in (1.37).

1.3.1.2. *Transverse Electric (TE) Waves.* Let $E_z = 0$. Then the electric vector is in the x, y plane, and H_z may be substituted for E_z in equation (1.31). Again the solutions are

$$\left.\begin{aligned}
X &= A\sin k_1 x + B\cos k_1 x\\
Y &= C\sin k_2 y + D\cos k_2 y
\end{aligned}\right\} \quad . \quad . \quad (1.38)$$

but in this case the boundary conditions take the form

$$E_x = 0 \text{ at } y = 0 \text{ or } b,\ E_y = 0 \text{ at } x = 0 \text{ or } a$$

Rewriting the scalar equations for this case,

$$\frac{Kj\omega}{c}E_x = \frac{\partial H_z}{\partial y} + \gamma H_y \quad . \quad (A)$$

$$\frac{Kj\omega}{c}E_y = -\gamma H_x - \frac{\partial H_z}{\partial x} \quad . \quad (B)$$

$$\frac{\partial H_y}{\partial x} = \frac{\partial H_x}{\partial y} \quad . \quad . \quad . \quad (C)$$

$$-\frac{\mu j\omega}{c}H_x = \gamma E_y \quad . \quad . \quad (D)$$

$$-\frac{\mu j\omega}{c}H_y = -\gamma E_x \quad . \quad . \quad (E)$$

$$-\frac{\mu j\omega}{c}H_z = \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \quad . \quad (F)$$

and substituting, the values for E_x, E_y, H_x, H_y in terms of H_z are obtained as follows—

$$k^2E_x = -\frac{\partial H_z}{\partial y}\frac{\mu j\omega}{c} \qquad H_x = -\frac{c\gamma}{\mu j\omega}E_y$$

$$k^2E_y = \frac{\partial H_z}{\partial x}\frac{\mu j\omega}{c} \qquad H_y = \frac{c\gamma}{\mu j\omega}E_x$$

Since E_x and E_y are respectively proportional to $\partial H_z/\partial y$ and $\partial H_z/\partial x$, the boundary conditions become

$$\left.\frac{\partial H_z}{\partial y}\right]_{y=0,b} = \left.\frac{\partial H_z}{\partial x}\right]_{x=0,a} = 0$$

and the sine term disappears from (1.38) giving

$$H_z = \cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-\gamma z} \quad . \quad . \quad . \quad (1.39)$$

and again, for transmission of energy, λ must be less than a certain critical value λ_c given by equation (1.36).

The explicit values for the field components in this case are—

$$E_x = j\frac{2\pi^2 n\mu}{b\lambda k^2}\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{-2\pi jz/\lambda_g}$$

$$E_y = -j\frac{2\pi^2 m\mu}{a\lambda k^2}\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-2\pi jz/\lambda_g}$$

$$E_z = 0$$

$$H_x = j\frac{2\pi^2 m}{a\lambda_g k^2}\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-2\pi jz/\lambda_g}$$

$$H_y = j\frac{2\pi^2 n}{b\lambda_g k^2}\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{-2\pi jz/\lambda_g}$$

$$H_z = \cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-2\pi jz/\lambda_g}$$

In the case of TE waves m or n may be zero, so the "dominant" wave is $TE_{0,1}$ or $TE_{1,0}$. These two are, of course, formally identical, since x and y are interchangeable.

Let $n = 0$. Then $E_x = H_y = 0$, and λ_c is given by

$$\lambda_c = 2a$$

As in the previous example, if $a = 4$ cm. and $b = 3$ cm., $\lambda_c = 8$ cm., and this is clearly the longest wave which can be transmitted without attenuation along the tube. This is one reason why the $TE_{1,0}$ wave is important. It is the mode which transmits the waves of greatest length, or, putting the matter in a more significant manner, if an oscillation of wave-length λ such that

$$8 > \lambda > 5 \text{ cm}.$$

is fed into the above-mentioned guide in the proper manner, then it will be propagated in the $TE_{1,0}$ mode only, since the cut-off wave-length for all the other modes is less than 5 cm.

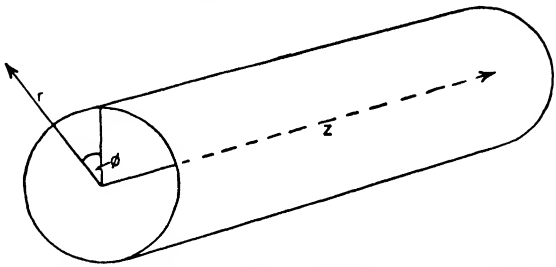


Fig. 6. A piece of cylindrical wave-guide showing co-ordinate axes

1.3.1.3. *Cylindrical Guides.* The technique employed for rectangular guides can be employed equally in the case of cylindrical guides if the scalar field equations are first transferred into cylindrical co-ordinates. Let the cylindrical guide be as shown in Figure 6, where the axis of the cylinder is the z-axis, and the other two co-ordinates are circular, r and ϕ.

Then the equations of (1.29) can be transformed, using

$$E_x = E_r \cos\phi - E_\phi \sin\phi$$
$$E_y = E_r \sin\phi + E_\phi \sin\phi$$

and $r = (x^2 + y^2)^{\frac{1}{2}}$ and $\tan\phi = y/x$.

They then become

$$\left.\begin{aligned}
\frac{K}{c}\frac{\partial E_r}{\partial t} &= \frac{1}{r}\frac{\partial H_z}{\partial \phi} - \frac{\partial H_\phi}{\partial z} && \text{(A)}\\
\frac{K}{c}\frac{\partial E_\phi}{\partial t} &= \frac{\partial H_r}{\partial z} - \frac{\partial H_z}{\partial r} && \text{(B)}\\
\frac{K}{c}\frac{\partial E_z}{\partial t} &= \frac{\partial H_\phi}{\partial r} - \frac{1}{r}\frac{\partial H_r}{\partial \phi} + \frac{1}{r}H_\phi && \text{(C)}\\
-\frac{\mu}{c}\frac{\partial H_r}{\partial t} &= \frac{1}{r}\frac{\partial E_z}{\partial \phi} - \frac{\partial E_\phi}{\partial z} && \text{(D)}\\
-\frac{\mu}{c}\frac{\partial H_\phi}{\partial t} &= \frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} && \text{(E)}\\
-\frac{\mu}{c}\frac{\partial H_z}{\partial t} &= \frac{\partial E_\phi}{\partial r} - \frac{1}{r}\frac{\partial E_r}{\partial \phi} + \frac{1}{r}E_\phi && \text{(F)}
\end{aligned}\right\} \quad (1.40)$$

As in the case of rectangular co-ordinates, differentiating one of these equations with respect to t and substituting from two others leads to the wave-equation.

For example,

$$-\frac{\mu}{c}\frac{\partial^2 H_z}{\partial t^2} = -\frac{c}{K}\left[\frac{\partial^2 H_z}{\partial r^2} + \frac{1}{r}\frac{\partial H_z}{\partial r} + \frac{1}{r^2}\frac{\partial^2 H_z}{\partial \phi^2} + \frac{\partial^2 H_z}{\partial z^2}\right]$$

$$+ \frac{c}{K}\frac{\partial}{\partial z}\left[\frac{\partial H_r}{\partial r} + \frac{1}{r}H_r + \frac{1}{r}\frac{\partial H_\phi}{\partial \phi} + \frac{\partial H_z}{\partial z}\right]$$

The terms within the second bracket make up the cylindrical transform of the quantity

$$\frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z}$$

and are therefore identically equal to zero in a non-conducting medium; whence the wave-equation is

$$\frac{\partial^2 H_z}{\partial r^2} + \frac{1}{r}\frac{\partial H_z}{\partial r} + \frac{1}{r^2}\frac{\partial^2 H_z}{\partial \phi^2} + \frac{\partial^2 H_z}{\partial z^2} - \frac{K\mu}{c^2}\frac{\partial^2 H_z}{\partial t^2} = 0 \quad . \quad . \quad (1.41)$$

Again, the typical component H_z may be written as a product

$$H_z = R\Phi Z e^{j\omega t}$$

and (1.41) becomes

$$\frac{1}{R}\frac{d^2R}{dr^2} + \frac{1}{rR}\frac{dR}{dr} + \frac{1}{r^2\Phi}\frac{d^2\Phi}{d\phi^2} + \frac{1}{Z}\frac{d^2Z}{dz^2} + \frac{K\mu\omega^2}{c^2} = 0$$

and if $Z = e^{-\gamma z}$

$$\frac{1}{R}\frac{d^2R}{dr^2} + \frac{1}{rR}\frac{dR}{dr} + \frac{1}{r^2\Phi}\frac{d^2\Phi}{d\phi^2} + k^2 = 0$$

where $$k^2 = \gamma^2 + K\mu\omega/c^2$$

As before, Φ may be written $A \sin k_1\phi + B \cos k_1\phi$, but in this case only one term need be used, since the choice of the radius $\phi = 0$ is arbitrary. Also, since Φ must have period $\phi = 2\pi$, k_1 must be an integer n, so the equation for R becomes

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} + k^2\left(1 - \frac{n^2}{k^2r^2}\right)R = 0$$

or, writing $kr = x$,

$$\frac{d^2R}{dx^2} + \frac{1}{x}\frac{dR}{dx} + \left(1 - \frac{n^2}{x^2}\right)R = 0$$

which is Bessel's equation.

Having regard to the fact that H_z must be finite when $r = 0$, Bessel Functions of the First Kind only need be used,

whence $$H_z = J_n(kr) \cos n\phi e^{-\gamma z} \quad . \quad . \quad . \quad (1.42)$$

Particular cases will now be discussed.

1.3.1.4. *Transverse Electric Waves.* Let $E_z = 0$. Then substituting in equations (1.40), the following values for the components are obtained, where $\gamma = j\beta = j \cdot 2\pi/\lambda_g$, and λ, λ_g are the wave-lengths in free space and the guide respectively.

$$\left.\begin{aligned} E_r &= j\frac{2\pi\mu n}{k^2\lambda r} J_n(kr) \sin(n\phi) e^{-2\pi jz/\lambda_g} \\ E_\phi &= j\frac{2\pi\mu}{k\lambda} J_n'(kr) \cos(n\phi) e^{-2\pi jz/\lambda_g} \\ E_z &= 0 \\ H_r &= -j\frac{2\pi}{k\lambda_g} J_n'(kr) \cos(n\phi) e^{-2\pi jz/\lambda_g} \\ H_\phi &= j\frac{2\pi n}{k^2\lambda_g r} J_n(kr) \sin(n\phi) e^{-2\pi jz/\lambda_g} \\ H_z &= J_n(kr) \cos(n\phi) e^{-2\pi jz/\lambda_g} \end{aligned}\right\} \quad . \quad . \quad . \quad (1.43)$$

Now writing in the boundary conditions, since $E_z = 0$, these are simply that $E_\phi = 0$ when $r = a$, and a = radius of the cylinder.

i.e. $$J_n'(ka) = 0 \quad . \quad . \quad . \quad . \quad (1.44)$$

This equation determines the permissible values of k to be used in the equation

$$k^2 = \gamma^2 + K\mu\omega^2/c^2$$

Let the values of ka obtained from equation (1.44) be associated with the integers 1, 2, 3, . . . in order of magnitude. Then the waves can be represented, as before, by $TE_{n,m}$ where m denotes the m^{th} root of the equation (1.44).

For example, if $n = 0$

$$J'_0(ka) = -J_1(ka) = 0 \text{ gives } ka = 3{\cdot}832 \ (m = 1),$$
$$7{\cdot}016 \ (m = 2),\ 10{\cdot}173 \ (m = 3), \text{ etc.}$$

The two most important practical cases are the $TE_{0,1}$ and the $TE_{1,1}$ waves for which $ka = 3{\cdot}832$ and $1{\cdot}841$ respectively. The critical wave-length below which transmission occurs is given by the equation

$$\frac{\mu K\omega^2}{c^2} = k^2_{n,m} \text{ or } \frac{4\pi^2}{\lambda_c^2} = k^2_{n,m}$$

i.e.
$$\lambda_c = \frac{2\pi}{k_{n,m}}$$

For a $TE_{0,1}$ wave, $\lambda_c = 2\pi a/3{\cdot}832$ and for a $TE_{1,1}$ wave $2\pi a/1{\cdot}841$.

The $TE_{0,m}$ type is a very interesting one. It will be noted that E_r and H_ϕ are zero in this case, the only finite components being E_ϕ, H_r and H_z. This combination of components causes the dissipation of energy in the wall (where this has a finite conductivity) to decrease as the frequency increases, a phenomenon which has no parallel in other forms of guided transmission. For this reason the $TE_{0,1}$ mode is very useful if a very high Q is required in a circuit.

1.3.1.5. *Transverse Magnetic Waves.* Let $H_z = 0$. Substituting as before, the following values for the components are obtained. Since the equations are symmetrical in H and E, if μ and K are interchanged, the values in this case can also be obtained from (1.43) by interchanging H for E and writing K for μ.

$$\left.\begin{aligned}
E_r &= -j\frac{2\pi}{k\lambda_g} J'(kr)\cos(n\phi)e^{-2\pi jz/\lambda_g} \\
E_\phi &= j\frac{2\pi n}{k^2\lambda_g r} J_n(kr)\sin(n\phi)e^{-2\pi jz/\lambda_g} \\
E_z &= J_n(kr)\cos(n\phi)e^{-2\pi jz/\lambda_g} \\
H_r &= j\frac{2\pi K n}{k^2\lambda r} J_n(kr)\sin(n\phi)e^{-2\pi jz/\lambda_g} \\
H_\phi &= j\frac{2\pi K}{k\lambda} J_n'(kr)\cos(n\phi)e^{-2\pi jz/\lambda_g} \\
H_z &= 0
\end{aligned}\right\} \quad . \quad . \quad . \quad (1.45)$$

The boundary condition is $E_z = 0$ at $r = a$, giving
$$J_n(ka) = 0 \quad . \quad . \quad . \quad (1.46)$$

As before, the $TM_{0,1}$ wave is identified with $n = 0$ and the first root of $J_0(ka) = 0$, and $TM_{1,1}$ with $n = 1$ and the first root of $J_1(ka) = 0$. The corresponding values of ka are 2·405 and 3·832. It will be noted that $TM_{1,1}$ gives the same value of k as $TE_{0,1}$, since $J_0' = -J_1$. For a $TM_{0,1}$ wave, $\lambda_c = 2\pi a/2{\cdot}405$ and for a $TM_{1,1}$ wave, $\lambda_c = 2\pi a/3{\cdot}832$. Of the dominant modes so far mentioned, $TE_{1,1}$ has the highest critical wave-length for a given diameter.

1.3.1.6. *Field Distribution in Wave-Guides.* The most important mode in a rectangular guide is the $TE_{1,0}$, while for cylindrical guides the $TE_{1,1}$, $TE_{0,1}$ and $TM_{0,1}$ are all important. In Figure 7 the direction of the electric and magnetic fields in the transverse plane are indicated for each of these modes. The plane chosen has been such that $E_z = 0$ in each case.

For the future guidance of the reader, the real components are evaluated for these modes, assuming the dielectric to be air in which μ and K are approximately unity.

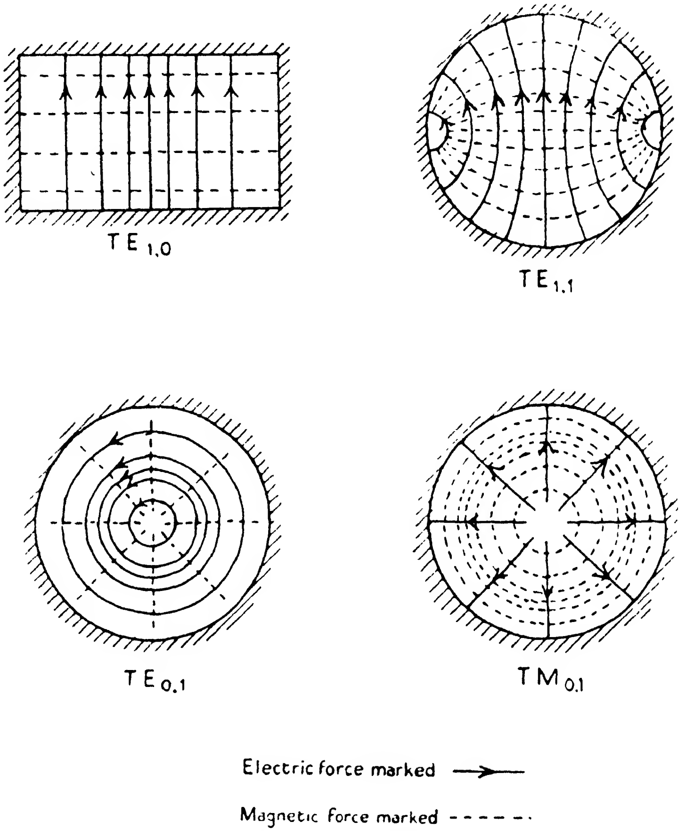


Fig. 7. Lines of electric and magnetic force in important dominant modes of propagation in metallic guides

(i) $TE_{1,0}$ Rectangular. Sides *a* and *b* parallel to *x* and *y* axis.

$$E_x = 0;\ E_y = \frac{2a}{\lambda} \sin\left(\frac{\pi x}{a}\right) \sin\left(\frac{2\pi z}{\lambda_g}\right);$$

$$H_y = 0;\ H_x = -\frac{2a}{\lambda_g} \sin\left(\frac{\pi x}{a}\right) \sin\left(\frac{2\pi z}{\lambda_g}\right);$$

$$H_z = -\cos\left(\frac{\pi x}{a}\right) \cos\left(\frac{2\pi z}{\lambda_g}\right)$$

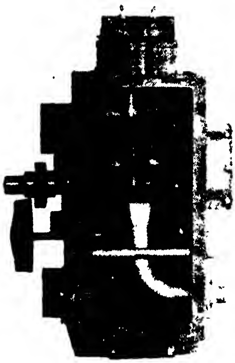


Fig. 1

Fig. 5

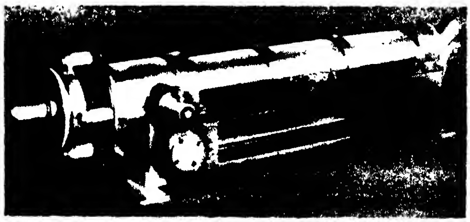


Fig. 2

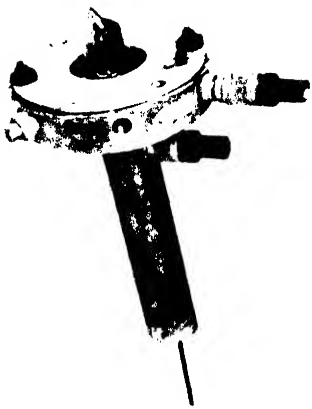


Fig. 4

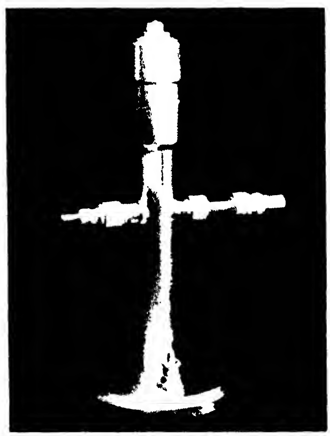


Fig. 3

CIRCUITS USED AT U.H.F.

Fig. 1. Valve with open line circuit.
Fig. 2. Valve with concentric line circuit.
Fig. 3. Concentric line wavemeter.

(ii) $TE_{1,1}$ Cylindrical. Radius *a*.

$$E_r = -\frac{2\pi a^2}{(1{\cdot}84)^2\lambda r} J_1\left(\frac{1{\cdot}84r}{a}\right)\sin\phi\sin\left(\frac{2\pi z}{\lambda_g}\right)$$

$$E_\phi = -\frac{2\pi a}{1{\cdot}84\lambda} J_1'\left(\frac{1{\cdot}84r}{a}\right)\cos\phi\sin\left(\frac{2\pi z}{\lambda_g}\right)$$

$$E_z = 0$$

$$H_r = \frac{2\pi a}{1{\cdot}84\lambda_g} J_1'\left(\frac{1{\cdot}84r}{a}\right)\cos\phi\sin\left(\frac{2\pi z}{\lambda_g}\right)$$

$$H_\phi = -\frac{2\pi a^2}{(1{\cdot}84)^2\lambda_g r} J_1\left(\frac{1{\cdot}84r}{a}\right)\sin\phi\sin\left(\frac{2\pi z}{\lambda_g}\right)$$

$$H_z = -J_1\left(\frac{1{\cdot}84r}{a}\right)\cos\phi\cos\left(\frac{2\pi z}{\lambda_g}\right)$$

(iii) $TE_{0,1}$ Cylindrical.

$$E_r = 0;\ E_\phi = \frac{2\pi a}{3{\cdot}83\lambda} J_1\left(\frac{3{\cdot}83r}{a}\right)\sin\left(\frac{2\pi z}{\lambda_g}\right)$$

$$E_z = 0;\ H_r = -\frac{2\pi a}{3{\cdot}83\lambda_g} J_1\left(\frac{3{\cdot}83r}{a}\right)\sin\left(\frac{2\pi z}{\lambda_g}\right)$$

$$H_\phi = 0;\ H_z = J_0\left(\frac{3{\cdot}83r}{a}\right)\cos\left(\frac{2\pi z}{\lambda_g}\right)$$

In this case E_ϕ and H_r are a maximum at $r/a = 0{\cdot}48$.

(iv) $TM_{0,1}$ Cylindrical.

$$E_r = -\frac{2\pi a}{2{\cdot}4\lambda_g} J_1\left(\frac{2{\cdot}4r}{a}\right)\sin\left(\frac{2\pi z}{\lambda_g}\right);\ E_\phi = 0$$

$$E_z = J_0\left(\frac{2{\cdot}4r}{a}\right)\cos\left(\frac{2\pi z}{\lambda_g}\right);\ E_r = 0$$

$$H_\phi = \frac{2\pi a}{2{\cdot}4\lambda} J_1\left(\frac{2{\cdot}4r}{a}\right)\sin\left(\frac{2\pi z}{\lambda_g}\right);\ H_z = 0$$

In this case E_r and H_ϕ are a maximum at $r/a = 0{\cdot}75$.

In all cases above the relation between λ and λ_g is given by

$$\frac{1}{\lambda_g} = \left[\frac{1}{\lambda^2} - \frac{k^2}{4\pi^2}\right]^{\frac{1}{2}}$$

where the values of *k* in the four cases treated are as follows—

(i) $k = \pi/a$, (ii) $k = 1{\cdot}84/a$, (iii) $k = 3{\cdot}83/a$, (iv) $k = 2{\cdot}40/a$.

1.3.1.7. *Resonators.* Imagine a guide which is terminated at both ends of its *z*-axis by metal plates. If the distance between the ends is an integral number of half wave-lengths (the wave-length being λ_g) then a system of standing waves will be set up by reflections at the metal boundaries on which $E_x = E_y = 0$, or $E_r = E_\phi = 0$. Such a box is a resonator in many respects similar to a resonating circuit consisting of a coil and condenser or to a half-wave concentric line. By attacking the problem in this manner the analysis of the preceding paragraphs may be extended to include resonators, although it would be formally more elegant and more general to give the x, y and z co-ordinates equal emphasis, and to write the solution of equation (1.30) symmetrically with respect to them. However, to avoid unnecessary repetition the resonator will be derived from waves travelling along the z-axis.

The system of standing waves is obtained by superimposing the wave travelling along the z-axis in a positive direction on a wave travelling along the axis in a negative direction. The second is obtained from the first by writing $-\gamma$ for γ throughout, or, since there is no attenuation and $\gamma = j\beta$, by writing $-\beta$ for β.

(i) As a first example, consider the TM modes of the rectangular guide. Then if the length along the z-axis is l,

$$l = p\lambda_g/2 \text{ or } \lambda_g = 2l/p \text{ with } p = 1, 2, 3, \ldots$$

and
$$\beta = \pi p/l$$

The standing wave pattern is given by

$$\left.\begin{aligned}
E_x &= \frac{\pi^2 pm}{alk^2} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) \sin\left(\frac{p\pi z}{l}\right) \\
E_y &= \frac{\pi^2 pn}{blk^2} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) \sin\left(\frac{p\pi z}{l}\right) \\
E_z &= \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) \cos\left(\frac{p\pi z}{l}\right) \\
H_x &= j\frac{2\pi^2 nK}{b\lambda k^2} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) \cos\left(\frac{p\pi z}{l}\right) \\
H_y &= -j\frac{2\pi^2 mK}{a\lambda k^2} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) \cos\left(\frac{p\pi z}{l}\right) \\
H_z &= 0
\end{aligned}\right\} \quad . \quad (1.47)$$

The pattern is distinguished by the values of the integers m, n, p and may be designated a $TM_{(m,n,p)}$ resonance. There are no waves corresponding to $(0, n, p)$ or $(m, 0, p)$, but the modes $(m, n, 0)$ for which $E_x \equiv E_y \equiv 0$ give a wave-pattern. Therefore the lowest real mode of the TM series is $TM_{(1,1,0)}$ the resonant frequency being determined by equation (1.35) with the substitutions $\lambda_g = \infty$, $m = n = 1$

or
$$f/v = \tfrac{1}{2}\left(\frac{1}{a^2} + \frac{1}{b^2}\right)^{\frac{1}{2}}$$

As a matter of fact it is not immediately obvious that p can be given the value zero in equation (1.47). As mentioned earlier, a rigid proof of this is only obtained by going back to the general field equations and inserting $\beta = 0$. This procedure shows that the resonant frequency of the $TM_{(1,1,0)}$ mode is independent of l.

(ii) The TE modes of the rectangular guide can be similarly treated. In this case the $TE_{(0,n,p)}$ and $TE_{(m,0,p)}$ waves exist, but p must be 1, 2, 3, . . . There is no case similar to $TM_{(m,n,0)}$. The mode of smallest frequency is $TE_{(0,1,1)}$ or $TE_{(1,0,1)}$ where

$$f/v = \tfrac{1}{2}\left\{\frac{1}{l^2} + \frac{1}{b^2}\right\}^{\frac{1}{2}}$$

(iii) The fundamental modes of the TM cylindrical series are $TM_{(0,1,0)}$ and $TM_{(1,1,0)}$ which are both independent of l as regards frequency. The frequencies are given by

$$\frac{f_{(0,1,0)}}{v} = \frac{2\cdot41}{2\pi a} \text{ and } \frac{f_{(1,1,0)}}{v} = \frac{3\cdot83}{2\pi a}$$

(iv) The fundamental modes of lowest frequency in the TE cylindrical category are the $TE_{(0,1,1)}$ and the $TE_{(1,1,1)}$, since waves do not exist for $p = 0$. The frequencies are given by

$$\frac{f_{(0,1,1)}}{v} = \tfrac{1}{2}\left\{\frac{1}{l^2} + \left(\frac{3\cdot83}{\pi a}\right)^2\right\}^{\frac{1}{2}}$$

and
$$\frac{f_{(1,1,1)}}{v} = \tfrac{1}{2}\left\{\frac{1}{l^2} + \left(\frac{1\cdot84}{\pi a}\right)^2\right\}^{\frac{1}{2}}$$

Throughout these derivations it should be observed that the wave velocity v of the wave in free dielectric is for a medium of constant K and permeability μ. This velocity should not be confused with the wave velocity in the guide. This is given by v_g, where

$$v_g = f\lambda_g = v\lambda_g/\lambda$$

It is quite possible that the wave-velocity in the guide should be greater than c the velocity in free space. The transfer of energy down the guide is to be associated with the "group" velocity or "signal" velocity which, as usual, is given by v_s, where

$$v_s v_g = v^2$$

1.3.2. *Practice.* As in the case of parallel wires and concentric lines, wave-guides can be used for many purposes—as circuits associated with valves to form oscillators or amplifiers, as wave-meters, or as transmission lines. A section of wave-guide may be used as a matching device in the manner already described for lines, and finally, by gradually increasing the cross-sectional area, an open guide may be made to radiate and so to act as an aerial. Such wave-guide "horns" are of great value, where a comparatively wide beam of micro-wave radiation is required.

1.3.2.1. *Amplifier and Oscillator Circuits.* As will be seen in the next chapter, it is difficult for many reasons to design valves suitable for use in the micro-wave region. Among these reasons is the fact that the inevitable inter-electrode capacities in the valve, and particularly the anode-grid capacitance of a triode, make it difficult to obtain a circuit of reasonable dimensions outside the valve envelope. But when a wave-guide is operated at a frequency higher than but adjacent to the cut-off value, the wave-length in the guide is large compared with the free space value. Hence a metal structure of reasonable dimensions can be constructed at frequencies at which concentric lines are too small. For example, suppose that a concentric line is used at 3000 Mc/s. Then a quarter-wave line is about 2½ cm. long, and the resonating length, if a valve is placed across the open end, will be much less. If, however, the $TE_{(1,0)}$ mode of a rectangular guide is used, the cut-off wave-length is $2a$ and the wave-length in the guide is given by

$$\lambda_a = \left[\frac{1}{100} - \frac{1}{4a^2}\right]^{\frac{1}{2}}$$

Let $a = 6$ cm. Then the guide will not transmit frequencies less than 2500 Mc/s, and the wave-length in the guide is about 18 cm. A quarter-wave section is therefore 4½ cm. long, comparing favourably with the concentric line length. A valve can be placed across such a guide, as shown in Figure 8, the guide acting as a resonant circuit between grid and anode. The valve is placed at a potential anti-node, although this is not essential. It can be tuned by moving the two sliding plungers A and B. A typical circuit is shown in Figure 4, Plate I. This is a complete unit, suitable for use as a small power amplifier or as a frequency multiplier in the 2700 to 3200 Mc/s region. Such wave-guide circuits possess an important advantage over all other circuits when used in this manner. They enhance the desired frequency and strongly attenuate all sub-harmonic frequencies. This has important applications in communication networks. For example, the circuit of Figure 4, Plate I, can be used as an amplifier or frequency multiplier with the plungers placed so that there are several half-waves in the guide. This could not be done, using a concentric line, since the latter would resonate on a sub-harmonic.

1.3.2.2. *High Pass Filters.* The last paragraph suggests incidentally that wave-guides may be used to cut off frequencies lower than that which is desired. A section of wave-guide introduced into a feeder line, or itself comprising the feeder, forms an excellent high pass filter. Moreover, the behaviour of such a filter may be accurately predicted from the equations of the previous section. The fact

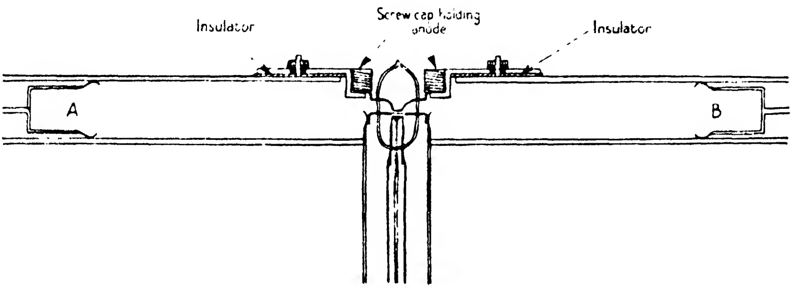


Fig. 8. A disc-seal valve with a rectangular cavity as the anode-grid circuit and a concentric line as the grid-cathode circuit

that attenuation was neglected does not affect the result appreciably unless in extreme cases.

1.3.2.3. *Wave-meters.* The fact that a closed section of metal tubing resonates at certain frequencies immediately suggests the use of such guides as wave-meters, and indeed, where the frequency band to be covered is relatively small, tube wave-meters are very efficient. Perhaps the best known of these, and certainly the most interesting, is the meter employing the $TE_{(0,1,1)}$ resonance in a cylindrical guide. As was mentioned at the end of paragraph 1.3.1.7 this mode is one in which the losses might be expected to be very small indeed, since the only finite force at the surface of the cylinder is H_z (see Figure 7). The resonant wave-length for a length l parallel to the axis is given by

$$\frac{1}{\lambda} = \frac{1}{2}\left\{\frac{1}{l^2} + \frac{(3\cdot 83)^2}{\pi^2 a^2}\right\}^{\frac{1}{2}}$$

and the cut-off wave-length by λ_c, where $\lambda_c = 2\pi a/3\cdot 83$. Suppose that a meter is required to measure wave-lengths in the neighbourhood of 3 cm. First, the radius of the cylinder is chosen to give a cut-off at a suitable wave-length above 3 cm., for example 3·5 cm.

This, as will be seen immediately, gives a fairly uniform scale at $\lambda = 3$ cm. Then $3{\cdot}83/2a = 4/7$, and

$$\frac{1}{\lambda} = \frac{1}{2}\left\{\frac{1}{l^2} + \frac{16}{49}\right\}^{\frac{1}{2}}$$

i.e. for $\lambda = 3$ cm., $l \fallingdotseq 3$ cm.

Also
$$\frac{d\lambda}{\lambda} = \frac{dl}{l} \cdot \frac{1}{1 + \dfrac{16l^2}{49}} \fallingdotseq \frac{1}{2}\frac{dl}{2l}$$

whence, if the wave-meter scale can be read to 0·001 cm., the wave-length can be measured to $1\frac{1}{2}$ parts in 10,000. This is only possible if the Q of the instrument is of the order of 10,000. The Q for this mode increases asymptotically as l increases to a theoretical value of about 100,000 for a silver cylinder. In practice a meter of the dimensions stated above can easily be made to have a working Q of the order of 20,000. One most important point must not be omitted. The $TE_{(0,1,1)}$ mode has a higher critical wave-length than others which may be excited, notably the $TM_{(0,1,0)}$ and the $TE_{(1,1,1)}$. It is therefore necessary to arrange to damp out these other oscillations without affecting the Q in the desired mode. Both the $TE_{(1,1,1)}$ and the $TM_{(0,1,0)}$ modes have radial components of electric force (see paragraphs 1.3.1.4 and 1.3.1.5), while the $TE_{(0,1,1)}$ mode has no radial or axial electric oscillation. In the resonator a series of wires are mounted, converging from the periphery of the cylinder at one end to a point on the axis some little way from the end. These "spokes" effectively short out the unwanted oscillations.* The second precaution which must be taken is to couple the power into and out of the resonator by means of probes aligned parallel to the cylindrical surface. These again assist in exciting only the wanted oscillation.

1.3.2.4. *Feeders.* The term "wave-guide" denotes one of the principal uses of metallic tubes, i.e. to transfer U.H.F. energy from one point to another. This is a problem which bristles with theoretical difficulties, and only a brief sketch of the outstanding phenomena can be given here. First, the guide should be chosen to transmit the wanted frequency with the minimum attenuation. Second, it is often important to choose the guide so that only one mode is transmitted. This can frequently be done, particularly if a rectangular guide is employed, for by suitably arranging the lengths a and b, the $TE_{(1,0)}$ mode can be made of much higher or lower frequency than the $TE_{(0,1)}$ mode. Third, it may sometimes be necessary to change the direction of the electric vector in the guide. This may be done in

* Alternatively, the moveable piston in the guide is of the non-contact type.

various ways, notably by inserting a cylindrical section with appropriate apertures. Finally, it is always necessary to match the guide to the generator and to the load, and, where different sections have been used, to match each section to the other. This calls for a knowledge of the "wave impedance" of each mode in each section. The concept of wave impedance has not been mentioned above.

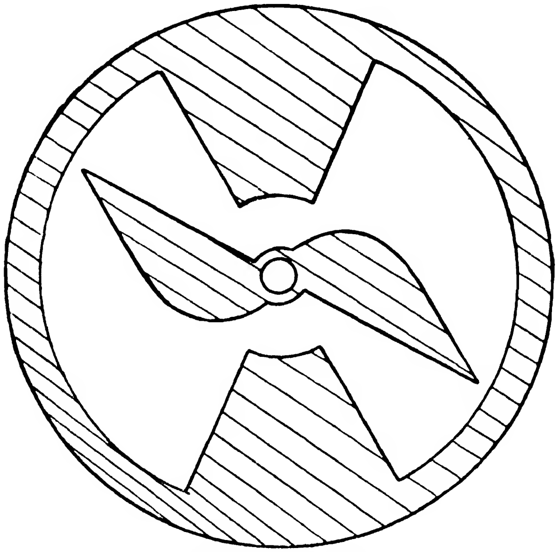


Fig. 9. Illustrating the principle of the "butterfly" circuit

Fortunately, the practical problem is not as formidable as theory would seem to suggest, for a satisfactory compromise is usually possible. Where the geometry does not lend itself to theoretical investigation, the judicious use of a standing wave detector will usually assist materially in solving the problem.

To give some idea of the order of magnitude of attenuation to be expected in micro-wave guides, a rectangular copper guide 3 in. by 1½ in. has an attenuation of the order of 10 db per 1000 ft. at 3000 Mc/s, when operated in the $TE_{(1,0)}$ mode. In all cases except one the attenuation in a guide increases more or less slowly as the frequency increases from the cut-off value. The exception is the $TE_{(0,m)}$ type of mode in the cylindrical guide. For such modes the attenuation asymptotes to zero as the frequency increases for reasons which have already been mentioned.

1.4. **Other U.H.F. Circuits.** As there is no end to the inventiveness of man, so there is no end to the variety of circuits which have been used or will be used at U.H.F. Some of the former partake of the properties of both distributed lines and lumped circuits, while some involve modifications of either which appear to be desirable from a mechanical viewpoint. One of the more important modifications is the so-called "butterfly" circuit, one form of which is shown diagrammatically in Figure 9.

The circuit consists essentially of a split stator condenser, the plates of which are connected together by a cylindrical inductance. This forms a resonant circuit. As the rotor is moved out from the stator, inductance and capacitance in the circuit simultaneously decrease, the former because the presence of the rotor arms partially "shorts" the flux linkage. The resonant frequency of the circuit can therefore be made to vary widely without serious alteration in the circuit shunt impedance, and such "butterflies" are well adapted for use in circuits where a wide frequency coverage is required. Practical oscillators have been made covering a 3 to 1 range, while absorption wave-meters which are not shunted by valves can be obtained covering a 4 to 1 range. Their maximum usable frequency is in the neighbourhood of 1000 Mc/s. Butterfly circuits cannot be made to have high Qs unless they are enclosed in a suitable metal box, for the structure has itself a high radiation resistance. This fact can occasionally be made use of, particularly where the butterfly is the resonant circuit of a receiver local oscillator, but, in general, the radiation is a trouble to the designer.

Other ingenious U.H.F. circuits will be mentioned from time to time in the subsequent chapters. It is probable that U.H.F. circuitry is in its infancy so far as radio communication is concerned.

REFERENCES

(1) Stratton, "Electromagnetic Theory," McGraw-Hill, 1941.
(2) Stratton, *loc. cit.*, p. 531-537.
(3) *Phil. Mag.*, May 1886.
(4) Willis Jackson, "High Frequency Transmission Lines," Methuen, 1945.
(5) Stratton, *loc. cit.*
(6) Huxley, "Principles and Practice of Wave Guides," Cambridge University Press, 1948.
(7) Huxley, *loc. cit.*
(8) Houstoun, "Treatise on Light," Longmans, 1938.

CHAPTER II

VALVES FOR USE AT ULTRA HIGH FREQUENCIES

2.1. **Limitations on the Use of Conventional Valves.** It has been said that the history of radio engineering is the history of thermionic devices, and this is a natural corollary from the fact that the valve is the only practical means whereby radio-frequency energy can be brought into being. In a transmitter the valves are the devices which convert the low frequency alternating currents of the supply into the radio-frequency alternating currents of the modulated carrier. In a receiver the valves, excited by the radio-frequency signal currents, amplify the latter at the expense of energy drawn from the supply. It is the inability of existing types of valves to perform this duty adequately at a certain part of the radio spectrum which automatically prevents the engineer from using any higher frequencies. It is the purpose of this section to describe the more important factors which cause conventional valves to fail, and to explain how the difficulties due to these factors may be partly overcome.

2.1.1. *Lead Impedance.* Even at moderately high frequencies (50 Mc/s) the impedance of valve leads begins to play an important part in the operation of the device. There are two reasons why this should be so.

(i) It becomes increasingly difficult to define the potential of a valve electrode, since it is impossible to get close enough to it with the desired circuit element. For example, it becomes impossible to "ground" a valve cathode, because the reactance of the cathode lead inside the envelope is no longer negligible.

(ii) As discussed below, the valve inter-electrode capacitances become increasingly important as the frequency increases. This causes larger radio-frequency currents to flow in the electrode leads. This in turn causes larger ohmic losses, unless the leads are specially prepared to have low resistance at high frequencies.

2.1.2. *Inter-electrode Capacitance and Input Impedance.* The grid-cathode capacitance of the average receiving valve is of the order of a few micro-microfarads. As the frequency increases, the associated reactance becomes small enough to determine the input impedance which in turn determines the power output of the valve. This in itself would limit the useful range of power valves, even if there were no other losses at the higher frequencies.

2.1.3. *Transit-time Effects.* The electrons in a valve take a finite time to traverse the distance between each pair of electrodes, but at frequencies less than a certain value which varies from valve to valve this finite time is very small compared with a period of the oscillation and can be neglected. It is on this hypothesis that conventional valve theory is built. When it becomes necessary to consider the effects of the time of transit of the electrons, the frequency is already so high that the other effects mentioned above can assume a dominant role. It is only when due allowance has been made for these other effects that a satisfactory transit-time theory can be attempted, and it cannot yet be said that this attempt has been wholly successful. There are many phenomena connected with the mechanism of space-charge limited emission in U.H.F. triodes which are not satisfactorily explained. Before discussing the matter further, therefore, it is necessary to describe the commonly used space-modulated U.H.F. valve in which the effects of lead inductance and resistance have been reduced to a minimum, and in which the electrode spacings are such that the transit-time effects can be neglected at frequencies very much greater than those employed with conventional valve types.

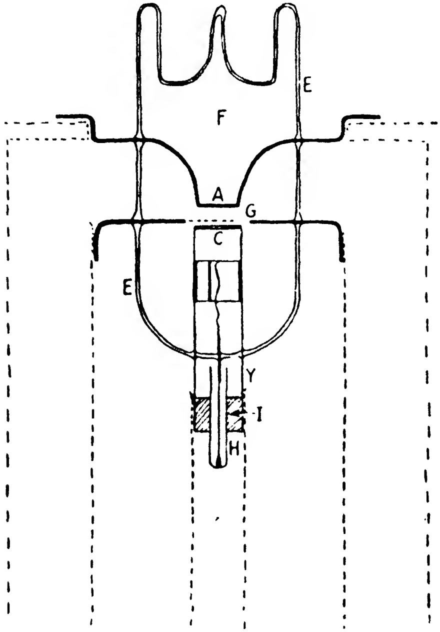


Fig. 10. Axial section of a typical disc-seal triode valve, the concentric lines being denoted by dashes

2.2. **The Disc-seal Triode.** Imagine a concentric line, closed at one end and open at the other. Such a device will act as a resonator at a wave-length given approximately by four times its length. The line would, therefore, form a suitable "tank" circuit for an amplifier or oscillator stage, if the valve could be connected across it at the free end with its grid to one cylinder and its anode to the other. This suggests a new type of valve construction, where the electrodes are brought out to a ring which can be fitted into a cylinder. This is achieved by means of the so-called "disc seal," which reduces lead inductance very considerably and simultaneously produces a construction geometrically suitable for line circuits. The idea must have occurred almost simultaneously to valve designers in U.S.A., Britain and Germany, for each country has produced a series of U.H.F. triodes of this type. A simple diagram, consisting of an axial section, of a typical British valve, the CV397, is shown in Figure 10. The diagram is not to scale. The intention is to bring out the essential points of construction only.

The valve has cylindrical symmetry, its anode A, grid G and cathode C being parallel discs. The cathode is rigidly connected to the cylinder Y which is directly sealed through the glass envelope E or connected by two or more short leads. The heater for the cathode is connected at one end to the latter. The free end is brought axially down the tube Y and soldered to a pin H. This pin is held axially with respect to Y by insulation I. The grid wires fill a central aperture in a dish-shaped copper stamping which is sealed directly into the envelope. The anode, another differently shaped stamping, is sealed into the glass in the same way. The getter flag is placed at F. It is the nature of the metal-glass seals which gives its name to the valve.

Such a valve is very suitable for use with concentric line circuits which may be arranged as shown by the dotted lines of Figure 10, one inside the other. In this way resonant circuits can be placed between cathode and grid and between grid and anode, care being taken to provide the necessary D.C. insulation. This procedure eliminates the difficulty of lead inductance referred to in paragraph 2.1.1 but it leaves the problem of the valve capacitances in a different form.

If the valve forms a shunt impedance $1/j\omega C$ at the end of a line of characteristic impedance Z_0, then by equation (1.22), the input impedance Z_I of a closed length l is given by

$$\frac{1}{Z_I} = j\omega C + \frac{1}{Z_0 \tanh \gamma l}$$

where $$\omega = 2\pi f = 6 \,.\, \pi \,.\, 10^{10}/\lambda$$

Let the losses in the line be negligible. Then $\gamma = j\beta$ and

$$\tanh \gamma l = j \tan \beta l = j \tan (2\pi l/\lambda)$$

Therefore

$$\frac{1}{Z_I} = j\left[\omega C - \frac{1}{Z_0 \tan (2\pi l/\lambda)}\right]$$

and the line is resonant at wave-length λ, if Z_I is infinite, i.e. when

$$\tan\left[\frac{2\pi l}{\lambda}\right] = \frac{\lambda}{6\pi 10^{10} Z_0 C} \quad . \quad . \quad . \quad . \qquad (2.1)$$

Figure 11 exhibits the mode of variation of l with λ, assuming $Z_0 = 70$ ohms and $C = 5\mu\mu F$. The value of C to be taken is *not* the valve inter-electrode capacitance, but the inter-electrode capacitance of the valve in the circuit. For a CV397, the anode-grid capacitance is seldom less than $5\mu\mu F$ under these circumstances.

The dotted line in Figure 11 is the corresponding $l_0 = \lambda/4$, showing

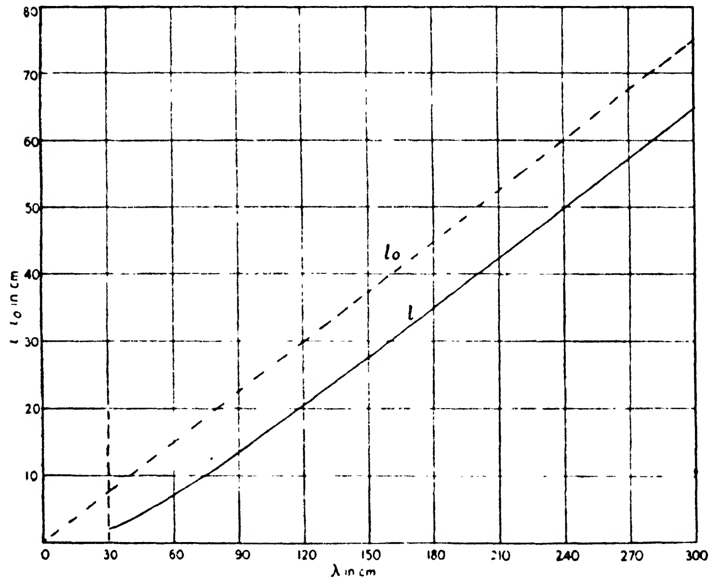
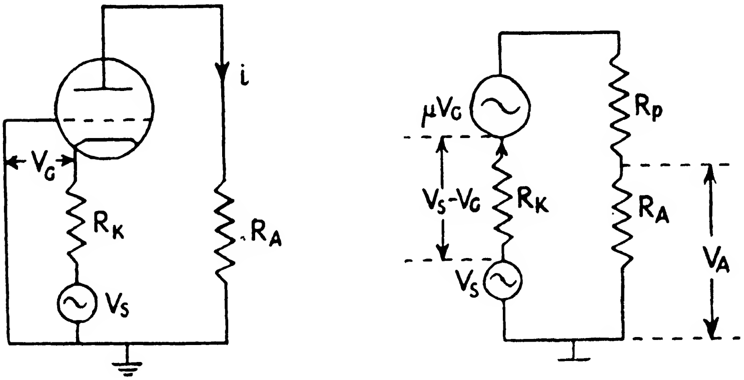


Fig. 11. Showing the effect of a lumped capacitance at the "open" end of a line in increasing its resonant wave-length

the shortening due to C. It is clear that for the figures mentioned a concentric line circuit could only be used at wave-lengths less than 30 cm. in a resonating mode higher than the quarter-wave.

It should be noted in passing that equation (2.1) can be applied equally to an open parallel wire line and may, indeed, be used to calculate the required characteristic impedance, when design calls for a certain length of line l at a given wave-length, as in paragraph 1.2.2.1.

As has already been mentioned, ring-seal or disc-seal valves have been made in U.S.A., Britain and Germany. Plate II shows photographs of some of the leading types. Figure 1 of the plate is the British CV397 and Figure 2 is a larger valve of the same type, the CV288. Figure 3 is one of the "lighthouse" series of American valves, corresponding roughly to the British CV397. Figure 4 is a war-time valve produced by Telefunken and named the LD6. It is between the CV397 and CV288 in size and performance. The British and American valves use metal-to-glass seals, while the German valves employed a ceramic in place of the glass. The intention in so doing was to obtain greater accuracy in electrode spacing than can be achieved with the same effort by metal-to-glass sealing. The ceramic used has a lower loss factor than the glass used in the British valves, but it is doubtful if this has any significance unless at the highest usable frequencies. It is almost standard practice now in



Figs. 12 and 13. The theoretical circuit of a grounded-grid triode amplifier and its equivalent

large U.H.F. triodes to adjust the cathode relative to the grid *after* manufacture. This gives greater uniformity and improved performance. Although all these valves have electrodes possessing cylindrical symmetry, and are therefore designed for use in coaxial circuits, they can be easily adapted to other line and wave-guide circuits. An example of this is shown in Figure 8, where a CV273 is exciting a wave-guide between grid and anode, the cathode-grid space being itself excited by a coaxial line. Once satisfactory contacts have been prepared, small valves of this type can be used in "butterfly" circuits also, although in this respect the CV273 and similar valves are to be preferred to valves having larger anode structures.

2.2.1. *Grounded-grid Operation of Disc-seal Triodes.* U.H.F. triodes, whether used as signal amplifiers, power amplifiers or oscillators are generally operated in some form of "grounded-grid" or "common grid" circuit. Such circuits are not, however, confined to use with disc-seal valves; indeed, the technique was invented before such valves existed. Devices of this type were sometimes called "inverted" amplifiers, and were used in association with conventional valves in power stages at frequencies of the order of 30 Mc/s.

The simple grounded-grid amplifier may be represented schematically by the circuit of Figure 12, where R_A is the load impedance (resistive) and R_K is the effective resistive impedance in series with the source. The equivalent circuit is shown in Figure 13, the valve being represented by the generator $-\mu V_g$.

Writing down the conditions in this circuit, $V_g = V_s - R_K i$, where V_s is the R.M.S. voltage at the source.

Also $$V_s - R_K i + \mu V_g - R_P i - R_A i = 0$$

Therefore $$i = \frac{V_s(\mu + 1)}{R_P + R_A + (\mu + 1)R_K}$$

and the voltage amplification is A_v, where

$$A_v = \frac{V_A}{V_s} = \frac{(\mu + 1)R_A}{R_P + R_A + (\mu + 1)R_K} \quad . \quad . \quad . \quad (2.2)$$

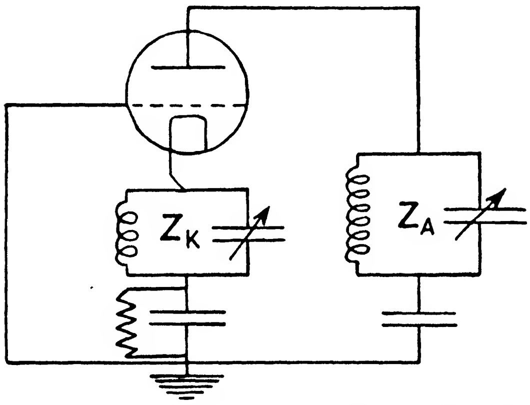


Fig. 14. The mode of operation of a grounded-grid amplifier

The equivalent circuit shows clearly the important features of this amplifier, (i) that the input is in series with the output, and (ii) that as a result of the finite impedance in series with the source, there is negative feed-back due to the common valve current. The grounded-grid amplifier is important in many connections at U.H.F. Certain aspects of its performance will be discussed in Chapter IV and again in Chapter V. So far as the present subject is concerned, an exact treatment cannot yet be given, since power operation involves non-linear characteristics. But the fundamental point with regard to the power amplifier and the frequency multiplier has already been mentioned, viz. the negative feed-back.

Figure 14 is a diagram of a typical radio-frequency grounded-grid stage.

This may be used for either voltage or power amplification. In either case the impedance Z_K should be kept as low as possible, particularly in a frequency multiplier. This tends to reduce the R_K of equation (2.2), and improves the voltage amplification automatically. So long as care is taken to see that the cathode circuit losses (external to the valve) are kept low, the reduction of R_K also improves the power amplification. The fact that Z_K should be small has a certain value from the circuit standpoint. Since a dynamic impedance of the order of 50–100 ohms is satisfactory, coil and condenser techniques can be used at frequencies as high as 400 Mc/s with consequent ease of tuning and small bulk. At frequencies above 400 Mc/s it is generally impossible to use a lumped circuit even in the cathode, and then recourse is made to lines in both cases. The details of such circuits will be considered in Chapter III.

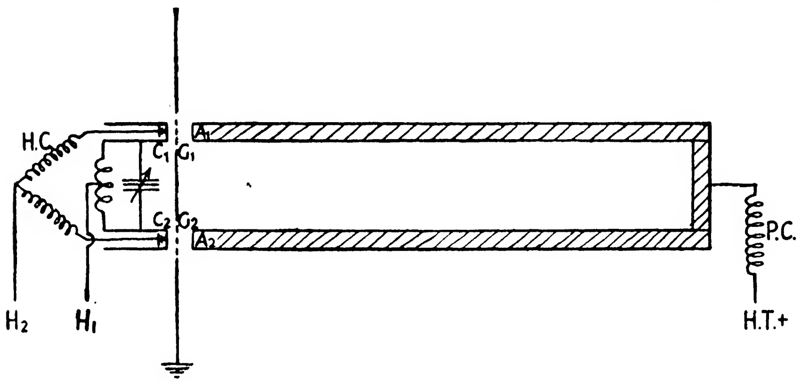


Fig. 15. An "open-line" circuit between the anodes of two grounded-grid valves operating as a push-pull amplifier

At the moment it is only necessary to establish the broad principles. In accordance with the theory above, it is desirable to use a line of low characteristic impedance for the input, and a line of high characteristic impedance for the output, but in practice many factors make it necessary to compromise on these requirements.

Grounded-grid operation is particularly well adapted to push-pull working, and at metric and decimetric wave-lengths parallel-wire lines are often employed. Figure 15 is a schematic of a push-pull grounded-grid amplifier for wave-lengths of the order of a metre.

Again, the input circuit is a balanced lumped circuit, the heater-cathode (positively biased) being connected at H_1 and the other heater connection being made at the electrical mid-point H_2 of a choke, marked H.C. The grids are directly grounded by being fitted into a metal screen, and the anode line circuit is connected to the plate supply through the choke P.C. at its mid-point. Valves of the CV273 and CV397 type are very suitable for use in such stages.

At still higher frequencies wave-guide circuits can be used in which the grid is common and may be regarded as grounded.

Figure 8 is an example of such a device. By using a suitably dimensioned wave-guide the cathode impedance can be kept as low as is desired.

2.2.2. *Double Concentric Line Circuits.* At frequencies of the order of 1000 Mc/s disc-seal valves are commonly used with concentric line circuits for input and output, the two lines being built coaxially. The diagram of such a circuit is shown in Figure 16, where A is the anode, G the grid, and C the cathode of the valve. S_A and S_G are capacitative plungers which effectively short-circuit the U.H.F. currents in their respective lines, but which enable the

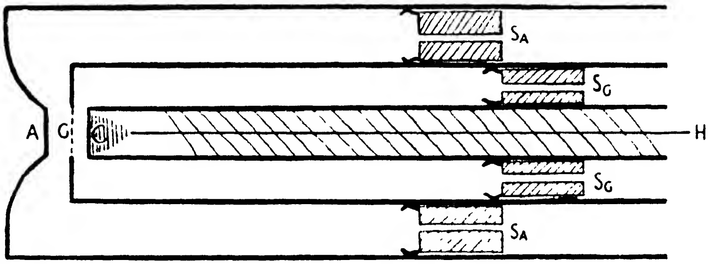


Fig. 16. A concentric line common-grid amplifier unit tuned by annular condensers which effectively short-circuit the radio frequency currents

three valve electrodes to be maintained at their correct D.C. operating potentials. How this can be done in practice will be explained in Chapter III. The valve heater lead is brought down the cathode cylinder to H. When the unit is used as an amplifier, the power is injected into the cathode circuit either capacitatively, by means of a probe, or inductively, by means of a loop. This excites the resonator formed by the inner concentric line terminated by the grid-cathode impedance of the valve. The output circuit is the outer concentric line, terminated by the anode-grid impedance of the valve. As described earlier (see equation (2.1)), the appropriate lines can be determined from a knowledge of the effective valve inter-electrode capacitances.

The equivalent lumped circuits are shown diagrammatically in Figure 17, and this unit is, in fact, the grounded-grid amplifier of Figure 14 except that the grid is common to both circuits but not necessarily earthed. Hence the generalized "grounded-grid" amplifier is represented by Figure 18, and the corresponding equivalent circuit is that shown in Figure 13 with the earthing point removed.

2.2.3. *Butterfly Circuits.* The butterfly circuit in its conventional form is not suitable for use single-ended and grounded grid. But if the circuit is regarded as being balanced with respect to the rotor

PLATE II

Fig. 1

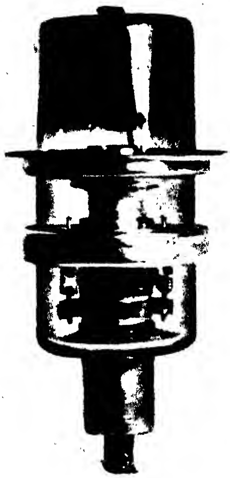


Fig. 2

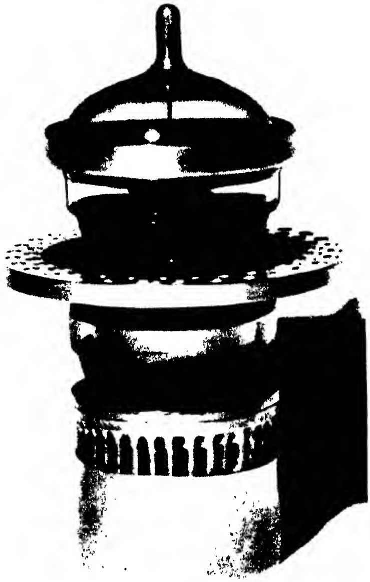


Fig. 3

Fig. 4

TYPICAL U.H.F. TRIODE VALVES

Fig. 1. British small disc-seal valve.
Fig. 2. British large disc-seal valve.
Fig. 3. American "lighthouse" valve.
Fig. 4. German ceramic disc-seal valve.

which is grounded, then two disc-seal valves may be attached to the appropriate symmetrical areas of the stator to form a push-pull amplifier with the grids grounded. The amplifier is formally identical with that of Figure 15, the butterfly taking the place of the line circuit. The cathode circuit in this case may be another butterfly of lower dynamic impedance, or it may be a conventional lumped circuit. In either case, the anode and cathode circuits can be conveniently ganged by means of the earthed spindles of the two rotors,

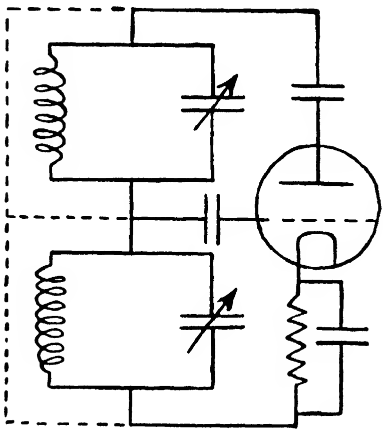


Fig. 17. The lumped circuit equivalent of Fig. 16

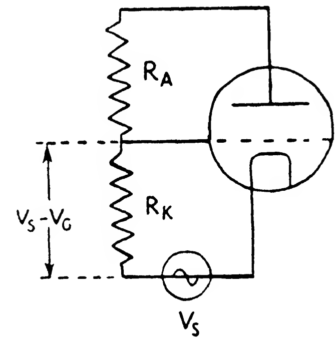


Fig. 18. The theoretical equivalent of Fig. 17

passing through the grounded screen. A specially constructed butterfly circuit, employing a CV273 valve and suitable for use as a receiver amplifying stage is shown in Plate I, Figure 5.

2.3. **Transit-time Effects.** The development of the disc-seal valve with parallel plane electrodes has removed the more important circuital limitations on the use of space-modulated triodes at ultra high frequencies, but it has done little to overcome the difficulties associated with the electron stream itself. Transit-time effects continue to be the most serious drawback to the use of space-modulated valves, and it is doubly unfortunate that the theory of these electron inertia effects bristles with difficulties which have been only partially overcome. The pioneer work in this field owes much to Benham[1] and, more recently, to Llewellyn[2] and Müller.[3] But although Llewellyn in particular has made a very brave effort to evolve a theory suitable for practical application, it must be admitted that the situation is still obscure. Much of the difficulty comes from the fact that the transit-time effects are so closely bound up with varying space-charge between the valve electrodes. Enough is not yet known of the behaviour of space-charge at ultra high frequencies. In the paragraphs which follow some of the more tractable cases of transit-time phenomena will be discussed, and it will be shown how the failure of space-modulation at the higher frequencies leads directly to the use of the velocity-modulation principle. So far as the communication engineer is concerned, the problem is to decide at what part of the spectrum space-modulation becomes uneconomic. The answer to this depends largely on the precision with which ultra high frequency valves are made, for it is still true that lack of precision in the manufacture and positioning of valve electrodes is (due to transit-time) the limiting factor in operation.

2.3.1. *Cathode Emission in a Space-charge Limited Planar Diode.* Before embarking on an analysis of transit-time effects, it is interesting to focus our attention on the role of cathode emission in high frequency valves.

Let the planar diode be represented by a cathode surface $x = 0$ at ground potential and a parallel anode surface at $x = d$ at potential V_d. Edge effects will be neglected.

Let I_0 be the current density in the valve per square centimetre, given by

$$I_0 = I_C + I_D$$

where I_C is the conduction current density, due to electron flow, and I_D is the displacement current density due to the valve reactance.

Then

$$I_0 = \rho u + \frac{10^9}{4\pi c^2}\frac{\partial E}{\partial t} \quad . \quad . \quad . \quad (2.3)$$

where ρ is the electron density in coulombs/sq. cm., u is the electron velocity in cm./sec., assumed parallel to the x-axis, c is the velocity of light in cm./sec., I_0 is in amp./sq. cm., and E is in volts/cm. This equation is the unidimensional form of the first equation of the electromagnetic field. ρ and u are both functions of x and t, but I_0 is (by definition) independent of x. Then the average value of E with respect to x is $E = V_d/d$

and

$$I_0 = (\rho u)_a + \frac{10^9}{c^2}\cdot\frac{1}{4\pi d}\frac{\partial V_d}{\partial t}$$

where $(\rho u)_a$ is the value of the conduction current corresponding to the average value of E. This shows clearly the capacitative nature of the displacement current, for $1/4\pi d$ is the "cold" capacity/sq. cm. of the diode.

Let

$$V_d = V_0 + V_A \sin \omega t$$

then

$$\frac{\partial V_d}{\partial t} = \omega V_A \cos \omega t$$

and

$$I_0 = (\rho u)_a + \frac{10^9}{c^2}\frac{V_A \omega}{4\pi d}\cos \omega t \quad . \quad . \quad . \quad (2.4)$$

Now in a space-charge limited diode at U.H.F. the time of flight of the electrons causes the "capacitative" current to have a certain component in phase with the voltage. Also, since this current varies as the frequency, the magnitude may be considerable. Hence, the cathode of a U.H.F. triode must be capable of giving a higher current density than a corresponding device operating at a lower frequency. Calculation shows that a diode which at low frequencies must produce a current density of 1 mA/sq. cm. to remain space-charge limited, must, under similar conditions of voltage, emit at the rate of 2 mA/sq. cm. at a frequency of 500 Mc/s. This in itself, and quite apart from other transit-time limitations, causes the design of U.H.F. valves to conform to a more stringent specification than in the case of their low frequency counterparts. There is no doubt that power amplifiers of the grounded-grid type tend to be temperature limited at frequencies considerably below that for which their design is otherwise suitable.

2.3.2. *Calculation of Transit-time in a Space-charge Limited Diode, the Anode Potential being Constant.* Returning to equation (2.3), insert the value of ρ given by the equation

$$\frac{10^9}{4\pi c^2}\frac{\partial E}{\partial x} = \rho$$

which expresses the conservation of charge in a unidimensional flow.

Then, since $u = dx/dt$,

$$I_0 = \frac{10^9}{4\pi c^2}\left[\frac{\partial E}{\partial x}\frac{\partial x}{dt} + \frac{\partial E}{\partial t}\right]$$

The expression within the brackets is the complete differential of E with respect to t, hence

$$I_0 = \frac{10^9}{4\pi c^2}\frac{dE}{dt} = a\frac{dE}{dt},$$

where a is the constant. In the steady state (V_d = constant), I_0 is a constant, and therefore dE/dt is a constant.

Now consider the movement of an electron in the field

$$\frac{d^2x}{dt^2} = \frac{e}{m}10^8E = 1{\cdot}77 \,.\, 10^{15}E = bE$$

Therefore

$$\frac{d^3x}{dt^3} = \frac{b}{a}I_0 = 2.10^{28}I_0$$

whence

$$\left.\begin{aligned} \frac{d^2x}{dt^2} &= 2.10^{28}\mathrm{I}_0 t \\ \frac{dx}{dt} &= 10^{28}\mathrm{I}_0 t^2 \\ x &= \frac{10^{28}}{3}\mathrm{I}_0 t^3 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (2.5)$$

where

$$\frac{d^2x}{dt^2} = \frac{dx}{dt} = x = 0 \text{ when } t = 0.$$

This is an approximation, for the cathode emits electrons with velocities distributed roughly according to the Maxwell law at a temperature corresponding to that in the vicinity of the emitter. For a temperature of 1000° Absolute, the electronic energies are distributed about a maximum at approximately $\frac{1}{16}$ of a volt, so the approximation is reasonably safe, so long as V_d is large compared with $\frac{1}{16}$ volt. Now let $x = d$ in the equations (2.5) above, and let $t = \mathrm{T}$, the transit-time and $dx/dt = v_d$, the terminal electronic velocity.

Then

$$v_d = 10^{28}\mathrm{I}_0\mathrm{T}^2$$

But

$$v_d = \sqrt{2 \,.\, \frac{e}{m}\, 10^8\mathrm{V}_d}$$

by the law of conservation of energy

$$= 5{\cdot}95 \,.\, 10^7\mathrm{V}_d^{\frac{1}{2}}$$

Therefore

$$\mathrm{T} = \frac{\sqrt{5{\cdot}95 \,.\, 10^7\mathrm{V}_d^{\frac{1}{2}}}}{10^{14} \,.\, \mathrm{I}_0^{\frac{1}{2}}} \quad . \quad . \quad . \quad (2.6)$$

Now substitute in the third equation of group (2.5)

$$d^2 = \frac{10^{56}\mathrm{I}_0^2(5{\cdot}95 \,.\, 10^7)^3\,\mathrm{V}_d^{\frac{3}{2}}}{9 \,.\, 10^{84} \,.\, \mathrm{I}_0^3}$$

or

$$\mathrm{I}_0 = 2{\cdot}34 \,.\, 10^{-6}\frac{\mathrm{V}_d^{\frac{3}{2}}}{d^2} \quad . \quad . \quad . \quad . \quad . \quad (2.7)$$

This is the well-known Child Law for a space-charge limited diode. Substituting for I_0 in equation (2.6),

$$\mathrm{T} = 5{\cdot}04 \,.\, 10^{-8}\frac{d}{\mathrm{V}_d^{\frac{1}{2}}} \quad . \quad . \quad . \quad (2.8)$$

For a temperature limited diode, where space-charge can be neglected, the corresponding transit-time T′ is given by

$$T' = \frac{2d}{\sqrt{\{2\,.\,1{\cdot}77\,.\,10^{15}\}}\,V_d^{\frac{1}{2}}}$$

$$= 3{\cdot}36\,.\,10^{-8}\frac{d}{V_d^{\frac{1}{2}}} = \tfrac{2}{3}T$$

Hence the time of transit in a space-charge limited diode is 1·5 times the corresponding time for a temperature limited diode.

2.3.3. *The Oscillating Diode.* Consider a plane parallel structure, consisting of a thermionic cathode C (Figure 19), a grid G and an anode A. Let there be a constant positive potential V_0 on the grid G, the cathode being grounded, and let there be an alternating potential difference $V_A \sin (pt + \phi)$ between G and A. Let the distance between G and A be D cm.

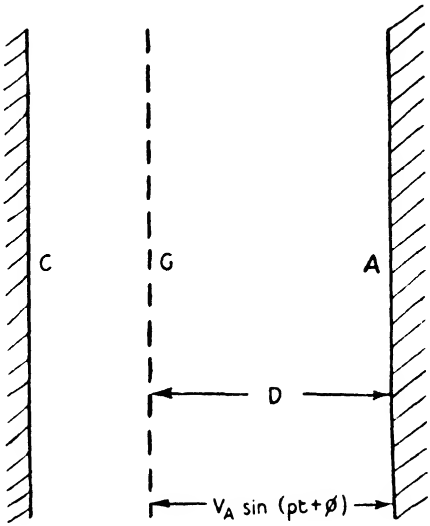


Fig. 19. A planar triode in which the grid is employed to accelerate the electrons and only a radio frequency field exists between the grid and the anode

Consider an electron in the space between G and A. Then, since there is no electron emission in the space, and since the electrons are moving with a finite speed due to V_0, space-charge effects are small, and the acceleration of the typical electron is given by

$$\frac{d^2x}{dt^2} = \frac{e}{m}\cdot\frac{V_A}{D}\sin (pt + \phi)$$

whence, integrating twice and noting that $dx/dt = \sqrt{2V_0 e/m}$ when $x = 0$ and $t = 0$,

$$\frac{dx}{dt} = \sqrt{2V_0 e/m} + \frac{e}{m}\cdot\frac{V_A}{pD}\left[\cos\phi - \cos (pt + \phi)\right] \quad . \quad . \quad (2.9)$$

and

$$x = \sqrt{2V_0 e/m}\,t + \frac{e}{m}\frac{V_A}{pD}\left[t\cos\phi + \frac{1}{p}\{\sin\phi - \sin (pt + \phi)\}\right] \quad (2.10)$$

Let $x = \mathrm{D}$ at $t = \mathrm{T}$, and write

$$\mathrm{T} = \frac{\mathrm{D}}{\sqrt{2\mathrm{V}_0 e/m}} + \delta$$

so that δ is the variation in the transit angle caused by the alternating field. Let $\mathrm{V_A}/\mathrm{V}_0$ be small. Then terms in δ^2, δ^3, . . . can be neglected. Let $\sqrt{2\mathrm{V}_0 e/m} = v_0$ and $e\mathrm{V_A}/mp\mathrm{D} = v_\mathrm{A}$ = amplitude of the additional electronic velocity due to the alternating field.

Then, substituting in equation (2.10),

$$\mathrm{D} = \mathrm{D} + v_0\delta + v_\mathrm{A}\left(\frac{\mathrm{D}}{v_0} + \delta\right)\cos\phi + \frac{v_\mathrm{A}}{p}\sin\phi - \frac{v_\mathrm{A}}{p}\sin\left(\frac{p\mathrm{D}}{v_0} + p\delta + \phi\right). \quad . \quad . \quad (2.11)$$

The last term on the right-hand side may be written

$$\sin\left(\frac{p\mathrm{D}}{v_0} + \phi\right) + p\delta\cos\left(\frac{p\mathrm{D}}{v_0} + \phi\right)$$

to the correct order of approximation.

Solving equation (2.11) for δ,

$$\delta = \frac{v_\mathrm{A}\left[\sin\left(\frac{p\mathrm{D}}{v_0} + \phi\right) - \sin\phi - \frac{p\mathrm{D}}{v_0}\cos\phi\right]}{p\left[v_0 + v_\mathrm{A}\left\{\cos\phi - \cos\left(\frac{p\mathrm{D}}{v_0} + \phi\right)\right\}\right]}$$

Substitute this value of δ in equation (2.9). Then

$$\left.\frac{dx}{dt}\right]_{t=T} = v_0 + v_a \quad \cos\phi - \cos\left(\frac{p\mathrm{D}}{v_0} + \phi\right)\Big] + v_\mathrm{A}{}^2\sin\left(\frac{p\mathrm{D}}{v_0} + \phi\right)\left(\frac{p\delta}{v_\mathrm{A}}\right)$$

Let $\frac{dx}{dt} = u$ = real velocity of the electron at A.

Form $u^2/v_0{}^2$ from the above expression, neglecting $(v_\mathrm{A}/v_0)^4$ and gher terms, and writing $p\mathrm{D}/v_0 = \theta$ = average transit angle. Then

$$\left(\frac{u}{v_0}\right)^2 = 1 + 2\frac{v_\mathrm{A}}{v_0}\Big[\cos\phi - \cos(\phi + \theta)\Big] + \left(\frac{v_\mathrm{A}}{v_0}\right)^2[\cos^2\phi + \cos^2(\phi + \theta) - 2\cos\phi\cos(\phi + \theta) + 2\sin(\phi + \theta)\{\sin(\phi + \theta) - \sin\phi - \theta\cos\phi\}]$$

The average value of $(u/v_0)^2$ over the cycle of the alternating field is

$$\frac{1}{2\pi}\int_0^{2\pi}\left(\frac{u}{v_0}\right)^2 d\phi = \xi$$

Then
$$\xi = 1 + \left(\frac{v_A}{v_0}\right)^2 \frac{1}{2\pi}\int_0^{2\pi}[2 - \tfrac{1}{2}\cos 2(\phi + \theta) - 2\cos\theta - \theta\sin(2\phi + \theta) - \theta\sin\theta]\,d\phi$$

$$= 1 + \left(\frac{v_A}{v_0}\right)^2 [2(1 - \cos\theta) - \theta\sin\theta]$$

Now write $(v_A/v_0)^2$ in terms of V_A, V_0 and θ.

Then
$$\xi = 1 + \left(\frac{V_A}{2V_0}\right)^2 \cdot \frac{1}{\theta^2}[2(1 - \cos\theta) - \theta\sin\theta]$$

or,
$$\tfrac{1}{2}m\bar{u}^2 = V_0 e + eV_A\left(\frac{V_A}{4V_0}\right)\frac{\{2(1 - \cos\theta) - \theta\sin\theta\}}{\theta^2} \quad . \quad (2.12)$$

where $\bar{u}^2$ is the mean square value of the electronic velocity.

If $\frac{1}{2}m\bar{u}^2 > V_0e$, the electrons have gained energy from the alternating field, and the device dissipates high frequency energy. If, however, $\frac{1}{2}m\bar{u}^2 < V_0e$, the device is a source of high frequency energy. This happens when

$$F = \frac{2(1 - \cos\theta) - \theta\sin\theta}{\theta^2} \quad . \quad . \quad . \quad . \quad (2.13)$$

is negative.

The equation holds for all values of θ, so long as V_A/V_0 is small. For small values of θ, F takes the form

$$F = \frac{\theta^2}{12} - \frac{\theta^4}{180} + \ldots$$

$$= \frac{\theta^2}{12}\left(1 - \frac{\theta^2}{15} + \ldots\right)$$

Figure 20 is a graph of F as a function of θ computed from equation (2.13).

Another way of using equation (2.12) is to write it in the form

$$\tfrac{1}{2}m\bar{u}^2 - V_0e = \frac{eV_A^2}{4V_0}F$$

Now multiply each side by n, where n is the number of electrons striking the anode per second.

Then $n(\frac{1}{2}m\bar{u}^2 - V_0e)$ is the average kinetic energy/sec. acquired by the electrons and may be written equal to $V_A^2/2R$, where $V_A^2/2$ is the mean square value of $V_A\cos(pt + \phi)$ and R is the damping resistance of the grid-anode space. Also $ne = I =$ the beam current, and therefore

$$\frac{IV_A^2}{4V_0} F = \frac{V_A^2}{2R}$$

or

$$R = \frac{2V_0}{I} \cdot \frac{1}{F}$$

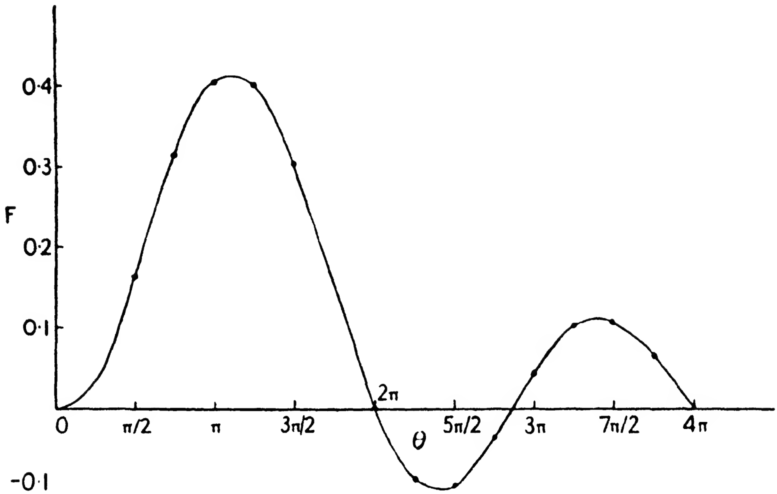


Fig. 20. The equivalent series resistance of the grid-anode space of the valve shown in Fig. 19 as a function of the transit angle

R is usually referred to as the beam damping impedance. It may have many values between $+\infty$ and $-\infty$, but is unimportant unless when it is small.

If the grid-anode space of the arrangement shown in Figure 19 is made part of a resonator, and if the potential difference V_0 is adjusted so that θ lies between 2π and $2{\cdot}86\pi$ (Figure 20), there is the possibility of oscillation occurring. The "negative resistance" of the equivalent diode may be small enough to overcome the resistive losses in the resonator. Oscillations of this type have been recorded by several observers,(4) but so far the energy output has been small. In some experiments made by the author in 1940 a few microwatts were obtained at a wave-length of 15 cm. This is in line with other results.

The real importance of equation (2.12) is that it shows how radically the behaviour of a valve may be altered by transit-time. Llewellyn(2) has made an analysis of the space-charge limited diode, and has shown that at all frequencies it behaves like a circuit element of resistance R and reactance X in series, where

$$R = 12R_0 \frac{2(1 - \cos\theta) - \theta\sin\theta}{\theta^4} \quad . \quad . \quad . \quad (2.14)$$

and
$$X = -\frac{1}{pC}\left[1 + \frac{\theta^3}{6}\{\theta(1 + \cos\theta) - 2\sin\theta\}\right]$$

The similarity between equations (2.12) and (2.14) will be noted.

In Llewellyn's equations R_0 is the usual diode static slope given by

$$R_0 = \frac{\partial V_d}{\partial I_0} = \frac{2}{3}\frac{V_d}{I_0}$$

and may be obtained directly from equation (2.7). The capacitance C is the "cold" value for the diode, $A/4\pi d$. It will be seen therefore that at any given frequency the diode behaves like a resistance and capacitance in series, but that the impedance does not vary in any simple fashion as the frequency is varied. The reactance approaches that of a condenser of capacitance $A/4\pi d$ as the frequency increases.

2.3.4. *The Input Impedance of the Negative Grid Triode.* The most important effect of transit-time in ultra high frequency communication circuits is on the input impedance of amplifier valves, and the associated phenomena have been the subject of a large number of theoretical and experimental investigations. Theoretically, the matter is summed up by Llewellyn[(2)] as follows: "However, it is known from experiment that the performance of negative grid tubes operated at such high frequencies and in their present environments is extremely poor. Under such circumstances the value of long computations needed to give the general performance curves is problematical." This is a rather remarkable comment on the present state of theory.

The fact is that all amplifiers exhibit a rapidly increasing active grid loss as the frequency increases. Much of this loss can be eliminated at frequencies less than 100 Mc/s by paying strict attention to stray grid, cathode and screen inductances, but at ultra high frequencies, using disc-seal valves the true transit-time phenomena appear. As in the case of the diode, the input impedance of the triode is a complicated function of the frequency. The resistive component can be positive, negative or zero, but again for a given frequency the impedance may be regarded as a resistance in series with a capacitance. The equivalent network suggested by Llewellyn is shown in Figure 21, where C is the cathode, G is the grid and P is the anode.

The fact that the grid resistances are in series with the grid capacitances gives rise to the well-known experimental result that the input power loss is proportional in many cases to the square of the frequency. It should be noted that R_{GP} of Figure 21 may be negative, but it is always unimportant compared with the loss due to R_{CG}.

C. N. Smyth[5] has recently called attention to the fact that when a diode is inserted across a resonant circuit at ultra high frequency

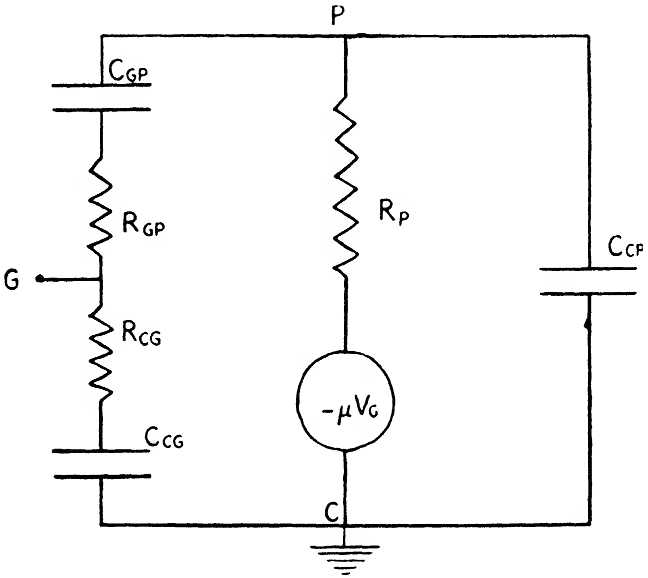


Fig. 21. The equivalent circuit of a triode valve according to Llewellyn

there is a damping which occurs when there is no appreciable direct current possible due to negative anode bias. The figures quoted by Smyth for a valve working at 3300 Mc/s with a spacing of 0·008 cm. are shown in Table IV. Smyth found that this damping depended

TABLE IV

Negative Bias (volt) . .	20	10	5	1
R/sq. cm. (ohm) . .	∞	Apprec.	200	50

upon cathode activity and could be reduced by cooling the cathode. He proposes the name "total emission damping" for the phenomenon. Clearly this damping must occur in the input circuit of a negative grid triode, and must be due to the finite velocity of emission of the electrons. The effect has been examined in more detail by A. van der Ziel,[6] his experiments being performed at a wave-length of 5·8 metres with a diode the spacing of which was 0·1 cm. A possible idealization of the conditions in the non-conducting diode has been examined by various writers,[7] and the theory is given in the next paragraph. The mathematical technique employed is similar to that used in paragraph 2.3.3.

2.3.5. *The Non-conducting Diode.* Let it be assumed that all the electrons are emitted from the cathode with velocity v_0, and that they then move under the action of a constant retarding field V_R/D superimposed upon a small alternating field $V_s \sin(\omega t + \phi)/D$. Then the equation of motion of the typical electron is

$$\frac{d^2x}{dt^2} = \frac{e}{m}\left[-\frac{V_R}{D} + \frac{V_s}{D}\sin(\omega t + \phi)\right]$$

giving

$$\frac{dx}{dt} = v_0 - \frac{e}{m}\frac{V_R}{D}\left[t + \frac{V_s}{\omega V_R}\left\{\cos(\omega t + \phi) + \frac{1}{\omega}\sin\phi\right\}\right]$$

and

$$x = v_0 t - \frac{e}{m}\frac{V_R}{D}\left[\tfrac{1}{2}t^2 - \frac{V_s}{\omega V_R}\left\{t\cos\phi - \frac{1}{\omega}\sin(\omega t + \phi) + \frac{1}{\omega}\sin\phi\right\}\right]$$

Consider the time taken by an electron to return to the cathode in the absence of the alternating field. Let this time be T_0.

Then $\quad T_0 = \dfrac{2v_0D}{V_R e/m}$ and $\omega T_0 = \theta =$ transit angle

Now let $T = T + \delta$ and substitute in the equation for x to find δ.

$$0 = -v_0\delta - \frac{e/m\,.\,V_R}{2D}\delta^2 + \frac{V_R e/m}{D}\frac{V_s}{\omega V_R}\left[T_0\cos\phi + \delta\cos\phi\right.$$
$$\left.-\frac{1}{\omega}\left\{\left(1 - \frac{\omega^2\delta^2}{2}\right)\sin(\omega T_0 + \phi) + \omega\delta\cos(\omega T_0 + \phi) - \sin\phi\right\}\right]$$

Multiply throughout by $\dfrac{2D\omega^2}{V_R e/m}$, and let $\dfrac{V_s}{V_R} = a$ where a is small.

Then

$$(\omega\delta)^2[1 - a\sin(\theta + \phi)] + \omega\delta[\theta - 2a\{\cos\phi - \cos(\theta + \phi)\}]$$
$$- 2a[\theta\cos\phi - \sin(\theta + \phi) + \sin\phi] = 0$$

or $$A(\omega\delta)^2 + B(\omega\delta) + C = 0$$

$$\begin{aligned}\text{Then } B^2 - 4AC &= \theta^2 - 4a[-\theta\cos\phi - \theta\cos(\theta + \phi)\\ &\quad + 2\sin(\theta + \phi) - 2\sin\phi]\\ &\quad + 4a^2[2 + \cos^2\phi - \cos^2(\theta + \phi)\\ &\quad - 2\cos\theta - 2\theta\sin(\theta + \phi)\cos\phi]\\ &= \theta^2 - 4aX + 4a^2Y = \theta^2\left[1 - \frac{4a}{\theta^2}(X - aY)\right]\end{aligned}$$

The term in the bracket can be expanded by the Binomial Theorem, since $X/\theta^2 \to a$ finite number when $\theta \to 0$.

$$\therefore \quad \sqrt{B^2 - 4AC} = \theta\left[1 - \frac{2a}{\theta^2}X + \frac{2a^2}{\theta^2}\left(Y - \frac{X^2}{\theta^2}\right)\right]$$

neglecting a^4, a^6, . . .

Now let $\cos\phi - \cos(\theta + \phi) = Z$

Then

$$\omega\delta = \frac{-B - \sqrt{(B^2 - 4AC)}}{2A}$$

$$= \frac{-\theta/2 + aZ + \theta/2 - aX/\theta + a^2(Y - X^2/\theta^2)/\theta}{1 - a\sin(\theta + \phi)}$$

To insert this value of $\omega\delta$ in the expression for dx/dt, rewrite

$$\frac{dx}{dt} = -v_0 - \frac{2v_0}{\theta}\omega\delta + \frac{2v_0}{\theta}\frac{V_s}{V_R}\{\cos\phi - \cos(\omega T_0 + \phi + \omega\delta)\}$$

Therefore, $$\frac{dx/dt}{v_0} = -1 + \frac{2aX}{\theta^2} - \frac{2a^2}{\theta^2}\left(Y - \frac{X^2}{\theta^2}\right)$$

Again writing the mean square value of $dx/dt = \bar{u}^2$,

$$\frac{\bar{u}^2}{v_0^2} = 1 + \frac{4a^2}{\theta^2}\int_0^{2\pi} Y d\phi$$

$$= 1 + \frac{4V_s^2}{V_R^2}\frac{1}{\theta^2}\{2(1 - \cos\theta) - \theta\sin\theta\}$$

As in the case of the conducting diode, write

$$\frac{\bar{u}^2}{v_0^2} = 1 + \frac{4V_s^2}{V_R^2}F,$$

whence $\frac{1}{2}m\bar{u}^2 - \frac{1}{2}mv_0^2 = \frac{4V_s^2}{V_R^2}V_0 e F$, where $V_0 e = \frac{1}{2}mv_0^2$,

and if N electrons are emitted by the cathode per second, the power taken from the alternating field is $V_s^2/2R$, where

$$\frac{V_s^2}{2R} = \frac{4V_s^2}{V_R^2}V_0 N e F$$

$$\text{or } \frac{1}{R} = \frac{8V_0 I}{V_R^2}F,$$

where I = total emission of the cathode. It should be noted that $V_0 I$ is the total power emitted by the cathode in the form of thermions.

The magnitude of $1/R$ given by the equation is adequate to explain Smyth's result, but, of course, the idealization made is not good enough to account for all the phenomena of total emission damping. As a second approximation to reality it might be supposed that θ varies more slowly than $1/V_R$ and that the distance D is a function both of V_R and of I.

2.4. Velocity Modulation. In a conventional valve amplifier the applied alternating potential triggers a current which by its variations produces alternating current power at the same frequency as the applied voltage. So long as the transit-time can be neglected, the current variations in the valve are related in a simple manner to the applied E.M.F. Such a valve amplifier may be said to operate on the principle of current or space-charge variation, or, alternatively, of space-charge modulation. This term is sometimes shortened to "space modulation." In the classical theory of the valve amplifier, it is quite unnecessary to know anything of the individual electron velocities. Apart from the matter of anode dissipation, these velocities might be regarded as infinite without prejudice to the theory.

As has been described in Section 2.3.3 the matter is very different at high frequencies. Here the electronic velocity becomes of major importance and classical valve theory fails because it takes no account of finite time of flight. The examination of the movements of the individual electrons under such circumstances has led to the discovery of a new principle of operation for amplifiers (and therefore oscillators) at these high frequencies. This may be described as a "velocity variation" or "velocity modulation" principle in contradistinction to the classical "space-charge variation" or "space-charge modulation" principle. It should be remarked at the outset that although the two principles are quite distinct, it is not always easy to tell whether a valve is operating on the one or the other. In some cases, as for example in the Barkhausen-Kurz or Gill-Morell modes of oscillation, both principles operate simultaneously in the one device. It is proposed in what follows to give an elementary description of the theory of velocity modulation devices, and then to describe in some detail the more important valves in use to-day.

2.4.1. *Theory of Velocity Modulation.* Imagine a beam of electrons moving at right angles to the set of parallel planes shown in Figure 22.

Let C be the thermionic emitter at ground potential, and let the grids A_1, A_2, A_3 and A_4 be maintained at a constant positive potential V_0. Now imagine an alternating potential $V_A \cos(pt + \phi)$ superimposed across A_1A_2. Then the electrons passing across the gap become velocity modulated as exhibited by equation (2.9), where D is the distance A_1A_2. As the electrons pass through A_2 they move into a force-free space between A_2 and A_3. Owing to the velocity variations produced in the gap A_1A_2, the electrons tend to produce condensations and rarefactions of charge density as they move towards A_3, and if the situation of A_3 is properly chosen, these condensations and rarefactions, crossing the gap A_3A_4, produce an alternating voltage across the electrodes by induction. This alternating voltage must obviously have the same frequency as the exciting voltage or a multiple of it, and so the device represents a method of transforming a voltage $V_A \cos pt$ at A_1A_2 into power at

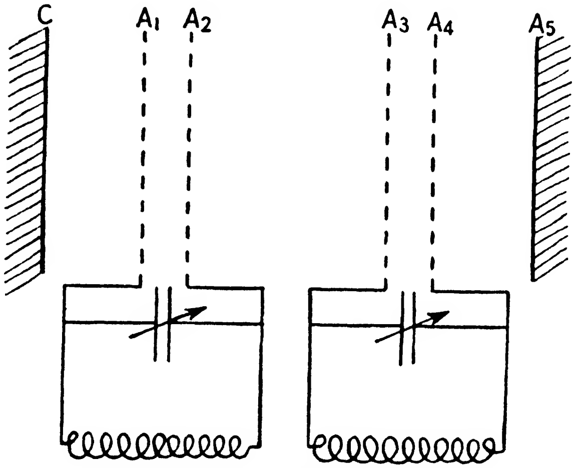


Fig. 22. A schematic representation of a klystron amplifier

the same frequency in a circuit attached to A_3A_4. Clearly the power will depend to some extent on the beam current, I_0, so, as in the conventional valve, there is a possibility of power amplification if energy can be abstracted from the steady source V_0. The electrons can be collected at the electrode A_5.

To complete the qualitative picture of the velocity modulation amplifier it is only necessary to imagine circuits such as those shown in Figure 22 attached to the two pairs of grids, so that the necessary transformations can be performed from input to output.

2.4.1.1. *Bunching*. As in the case of the conventional valve, it is convenient to start with certain simplifying assumptions. First, let it be supposed that the distance $D = A_1A_2$ is such that transit-time can be neglected. In fact, let A_1A_2 be negligibly small. Second, as usual, let the ratio V_A/V_0 be small so that higher powers of the ratio may be neglected. Let time be measured from some arbitrary value. Then the velocity of an electron after traversing A_1A_2 is given by v, where pt_1 is the arbitrary phase of the alternating voltage, and

$$\tfrac{1}{2}mv^2 = e(V_0 + V_A \sin pt_1)$$

or $$v = \sqrt{2V_0e/m}\left\{1 + \frac{V_A}{V_0}\sin pt\right\}^{\frac{1}{2}} \fallingdotseq v_0 + v_1 \sin pt_1$$

The motion of the electron in the space A_2A_3 will now be examined. This space is force-free, and therefore, if the distance $A_2A_3 = x$

$$x = (t_2 - t_1)(v_0 + v_1 \sin pt_1)$$

or $$t_2 \fallingdotseq t_1 + \frac{x}{v_0}\left(1 - \frac{v_1}{v_0}\sin pt_1\right) \quad . \quad . \quad . \quad (2.15)$$

where t_2 is the time at which the electron arrives at A_3.

Let the current at the gap A_1A_2 be I_0 (a constant) and let the current at A_3A_4 be I_2. Then if N electrons cross A_2 in time δt_1 and the same number cross A_3 in time δt_2,

$$\frac{I_2}{I_0} = \frac{\delta t_1}{\delta t_2} = \frac{dt_1}{dt_2} \fallingdotseq \frac{1}{1 - \frac{pxv_1}{v_0^2}\cos pt_1} \quad . \quad . \quad (2.16)$$

To examine the variation of the current I_2 due to velocity modulation at A_1A_2 it is necessary to write I_2 in terms of x and the transit angle. Let $pt_1 = \theta =$ the phase of the electron at A_1A_2, and let

$$\mu = p\left(t_2 - \frac{x}{v_0}\right), \text{ where } \frac{px}{v_0} = \text{transit angle across } A_2A_3$$

Also write $pxv_1/v_0^2 = a$. Then equation (2.15) becomes

$$\mu = \theta - a \sin \theta$$

and equation (2.16) becomes

$$I_2 = \frac{I_0}{1 - a\cos\theta}$$

Expand $1/(1 - a\cos\theta)$ as a Fourier series, using μ as the independent variable.

Then $$\frac{1}{1 - a\cos\theta} = 1 + \sum_{n=1}^{\infty} A_n \cos n\mu$$

and $$A_n = \frac{2}{\pi}\int_0^{\pi} \frac{\cos n\mu}{(1 - a\cos\theta)}\, d\mu$$

But $$d\mu = (1 - a\cos\theta)d\theta$$

$\therefore$ $$A_n = \frac{2}{\pi}\int_0^{\pi} \cos(n\theta - na\sin\theta)\, d\theta$$

This is Bessel's Integral. Its value is $\pi J_n(na)$.
Therefore

$$I_2 = I_0\,[1 + 2\{J_1(a)\cos\mu + J_2(2a)\cos 2\mu + J_3(3a)\cos 3\mu + \ldots\}]$$

Each harmonic term is a travelling wave of the form

$$J_n\left(\frac{npxv_1}{v_0^2}\right)\cos n\left(pt-\frac{px}{v_0}\right)$$

and so far as the first harmonic is concerned

$$I_2 = 2I_0\, J_1\left(\frac{pxv_1}{v_0^2}\right)\cos\left(pt-\frac{px}{v_0}\right) \quad . \quad . \quad (2.17)$$

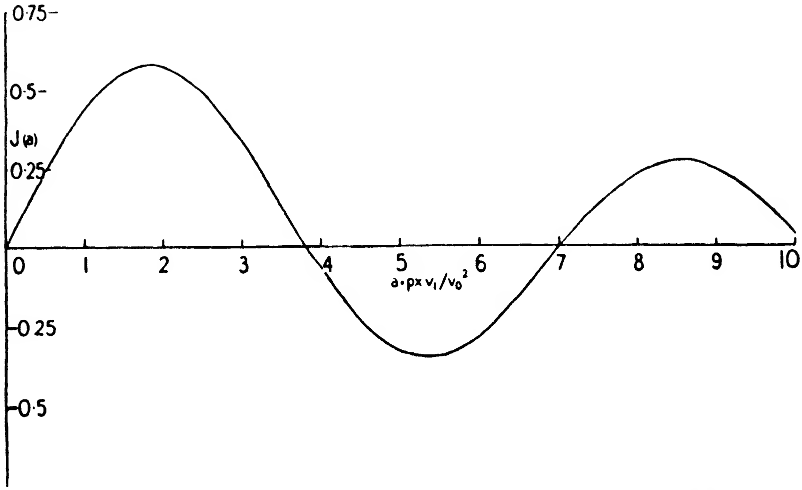


Fig. 23. J_1 (a) is a measure of the degree of bunching in the absence of space charge

It should be noted that the Fourier expansion only holds for

$$\frac{px}{v_0}\frac{v_1}{v_0} = \mu\frac{v_1}{v_0} < 1$$

Beyond this point it should be used with care. But equation (2.17) is substantially correct for all practical values of a.

The amplitude of the current term of the form $\cos(pt-\phi)$ is given by $J_1\left(\frac{pxv_1}{v_0^2}\right) = J_1(a)$ which is shown in Figure 23. The turning points occur at $a = 1{\cdot}84$, 5·33, 8·54 and 11·71.

This production of an alternating current from a direct current due to the "bunching" of the electrons is a direct result of the velocity modulation in the gap A_1A_2, the so-called "buncher" gap, and it is this alternating current in the beam flowing across A_3A_4 which produces by induction the alternating current power in the circuits across A_3A_4. The gap A_3A_4 is called the "catcher" gap, because it is there that the energy of the bunched electrons is caught by the circuit. From this simplified theory it would appear that the maximum current occurs at $a = 1{\cdot}84$ or at

$$x_B = \frac{1{\cdot}84\, v_0^2}{pv_1}$$

A more critical examination of the current as a function of θ shows that beyond this point x_B the electron bunches split into two which gradually move apart to coalesce with others at the second turning point of $J_1(a)$, namely at $a = 5{\cdot}33$. The maximum value of I_2 is given by

$$I_2 = 2\,.\,0{\cdot}58\, I_0 \cos(pt - \phi)$$

2.4.1.2. *Theoretical Efficiency of a Klystron Amplifier.* Assuming that all the high frequency energy at frequency $2\pi p$ can be extracted from the beam at the catcher gap, the output power will be

$$P = \frac{2\,.\,0{\cdot}58}{\sqrt{2}}\,.\,I_0\,.\,\frac{V_2}{\sqrt{2}}$$

where V_2 is the voltage across the catcher. For maximum power transfer $V_2 = V_0$, for if V_2 were greater than V_0 some of the electrons would be turned back to the cathode.

Therefore $$P = 0{\cdot}58\, I_0V_0$$

But the input direct-current power is $P_0 = I_0V_0$.

Therefore, the theoretical maximum klystron efficiency is 58%. Assuming that the ohmic losses in the catcher circuit absorb half the power (which will be the case with most practical values of I_0), then the available power is $0{\cdot}29\, I_0V_0$. Also, the transit-time in the gap A_1A_2 cannot in practice be neglected, so that the available power is lower still, leading to efficiencies of the order of 20% to 25%. If there is considerable loss of beam current at the grids A_1, A_2, A_3, this still further reduces the efficiency.

The best copper disc-seal klystrons made so far have had efficiencies of the order of 20%. It is improbable that this figure will be improved upon radically.

2.4.1.3. *The Coupling Coefficient.* One aspect of klystron theory, mentioned above, is worth a little more consideration. The time of transit of the electron across the gap A_1A_2 causes its velocity modulation to be less than that expressed by the equation

$$\tfrac{1}{2}mv^2 = e(V_0 + V_A \sin pt)$$

For, if $$\frac{d^2x}{dt^2} = \frac{e}{m}\frac{V_A}{D}\sin pt$$

then the electron starts across the gap at time $t - \frac{D}{2v_0}$ and emerges at time $t + \frac{D}{2v_0}$, and its velocity is

$$v = v_0 + \frac{e}{m}\frac{V_A}{D}\int_{t - D/2v_0}^{t + D/2v_0} \sin pt\, dt$$

$$= v_0 + \frac{e}{m}\frac{V_A}{pD}\left[\cos p\left(t - \frac{D}{2v_0}\right) - \cos p\left(t + \frac{D}{2v_0}\right)\right]$$

$$= v_0 + 2\frac{e}{m}\frac{V_A}{pD}\sin pt \sin\frac{pD}{2v_0}$$

Let the transit angle be θ, where $\theta = p\tau = pD/v_0$.

Then $$\tfrac{1}{2}mv^2 = \tfrac{1}{2}m\left(v_0^2 + 4\frac{e}{m}\frac{v_0 V_A}{pD}\sin\frac{pD}{2v_0}\sin pt\right)$$

to first order approximation $= e\left(V_0 + \frac{\sin\theta/2}{\theta/2}V_A \sin pt\right)$

Comparing this with the zero transit-time equation, it is seen that the effect of the finite width of the gap is to decrease the change in velocity of the electron by a factor $\beta = \frac{\sin\theta/2}{\theta/2}$. In fact

$$v_1 \doteqdot \frac{e}{m}\frac{V_A}{v_0}\beta \sin pt \quad . \quad . \quad . \quad (2.18)$$

β is known as the coupling factor between beam and circuit.

2.4.1.4. *Power Absorbed in Buncher Gap.* Returning to Section 2.3.3, it is shown there that the damping resistance of a space such as the buncher gap A_1A_2 is .

$$R = \frac{2V_0}{I_0}\frac{\theta^2}{2(1 - \cos\theta) - \theta\sin\theta}$$

This means that for all values of θ such as would occur in the buncher gap ($0 < \theta < \pi$), the beam absorbs energy from the input circuit. This damping resistance can be important if I_0 is large and θ approaches π.

2.4.1.5. *Space-charge Debunching.* For large values of I_0 the bunched beam is subject to forces due to its own charge density which tend to "debunch" it. These forces spread the beam bunches out in all directions and seriously affect the efficiency in power klystrons. The treatment is beyond the scope of this book. Reference should be made to the work of Webster[8] and others.

2.4.1.6. *Klystron Frequency Multipliers.* Owing to the form of the current pulse in the "drift" space between A_2 and A_3, namely

$$I_2/I_0 = 1 + 2 \sum_1^{\infty} J_n(na) \cos\left(npt - \frac{npx}{v_0}\right)$$

the beam current is extremely rich in harmonics, and this immediately suggests the use of the klystron as a frequency multiplier. Indeed, simple theory suggests that the efficiency of the klystron as a frequency trebler should be about 75% of the efficiency of the klystron as an amplifier. In practice the exact adjustment of a multiplier is tricky and it is even more difficult than in the case of the amplifier to obtain sufficiently high values of I_0. In spite of this, klystron multipliers are exceedingly promising for work in the 1000 to 5000 Mc/s region.

2.4.1.7. *The Klystron Oscillator.* If a device operates as an amplifier it can be made to oscillate by feeding back part of the energy in the proper phase. If, therefore, a fraction of the output from the catcher circuit of the klystron amplifier is fed back to the buncher circuit, conditions can be arranged such that the oscillations are self-maintained. An analysis of the klystron oscillator is a combination of the theory of velocity modulation and the theory of coupled circuits, and, as in the case of the triode oscillator, there is the possibility of two modes of oscillation. When the coupling is small, there is one mode only and one frequency: when the coupling is large, there are two modes of different frequency.

(*a*) *Small Coupling.* Referring back to Section 2.4.1.1, the bunched current is $2I_0J_1(a) \cos p(t - \tau)$ where τ is the transit-time in the drift space A_2A_3. This current is due to the electrons which crossed the gap A_1A_2 when the voltage across it was $V_1 = V_A \sin pt$. Therefore I_2 and V_1 differ in phase by $p\tau + \pi/2$. If the klystron oscillator is properly tuned, the impedance presented to the output is resistive and V_2 is in phase with I_2, whence the phase difference $p\tau + \pi/2$ exists between V_1 and V_2. But for loose coupling, the phase difference due to the coupling element is $\pi/2$ in order to maintain the frequency at the same value in both circuits. Therefore the total change of phase round the loop $A_1A_2 - A_3A_4 - A_1A_2$ is $p\tau + \pi$. For oscillation

$$p\tau + \pi = 2n\pi \text{ or } p\tau = \pi(2n - 1) \quad . \quad . \quad (2.19)$$

But $\tau = x/v_0$, where x is the drift length and $v_0 = \sqrt{2V_0e/m}$

$$\therefore \quad V_0 = \frac{p^2x^2}{2\pi^2(2n-1)^2e/m} = \frac{A}{(n-\frac{1}{2})^2} (n = 1, 2, \ldots) \quad . \quad (2.20)$$

for a given klystron.

This means that the under-coupled klystron oscillates when V_0 is in the vicinity of the discrete values given by (2.20).

(*b*) *Large Coupling.* If the two circuits are over-coupled, the phase difference due to the coupling element is 0 or π and the equation (2.19) becomes

$$p\tau + \pi/2 + (0 \text{ or } \pi) = 2n\pi$$

or

$$p\tau = \pi(2n \pm 1/2) \qquad . \quad . \quad . \ (2.21)$$

giving

$$V_0 = \frac{A}{(n \pm \frac{1}{4})^2} \qquad . \quad . \quad . \quad . \ (2.22)$$

It should be noted that the two modes, given by the + and − signs have different frequencies. It is astonishing how closely theory and experiment agree in this respect. In a klystron oscillator made by the author in 1941 the value of the drift space x calculated from equation (2.20) for different values of n was given by the following table—

V_0	820	980	1180	1470	1850
$n \pm \frac{1}{4}$	$6\frac{1}{4}$	$5\frac{3}{4}$	$5\frac{1}{4}$	$4\frac{3}{4}$	$4\frac{1}{4}$
x (cm.)	3·35	3·38	3·38	3·42	3·42

The measured value of x was 3·50 cm.

In this respect a word of warning is necessary. Although in the above theory the modes of oscillation have been associated with frequencies, the oscillations are maintained over a range of values of V_0 in the vicinity of the values given by equation (2.20) or (2.22). If V_0 differs from the values given, it means that the phase difference due to the coupling element differs from 0 or $\pi/2$ or π, as the case may be. This means a change in the frequency of oscillation, as predicted by the theory of coupled circuits, and therefore the frequency of the klystron oscillator changes as V_0 changes within the range of the latter which produces oscillations. To each value of V_0 there corresponds a particular frequency, showing that the klystron can be tuned over a small range by electronic methods. The frequency change is greater for an under-coupled oscillator than it is for an over-coupled oscillator, and this fact has important practical significance as will be seen later when the magnitude of the effect is discussed.

Many other factors influence the frequency of the klystron. Some of these will be discussed in the chapter on Frequency Control and some in the chapter on Modulation.

2.4.1.8. *Reflex Klystron Oscillators.* An ingenious modification of the velocity modulation principle makes it possible to construct an oscillator, using only one tuned circuit. Consider the arrangement shown in Figure 24 in which an electron beam, emitted by the cathode C is accelerated by the voltage V_0 to the gap A_1A_2 where it is velocity modulated by the signal voltage $V_A \sin pt$. Now imagine the beam to emerge into the retarding electric field produced by the presence of the reflecting electrode A_3 at the potential $- V_R$. The electrons advance into the space A_2A_3 with diminishing velocity and eventually turn back to the gap which they re-traverse in the opposite

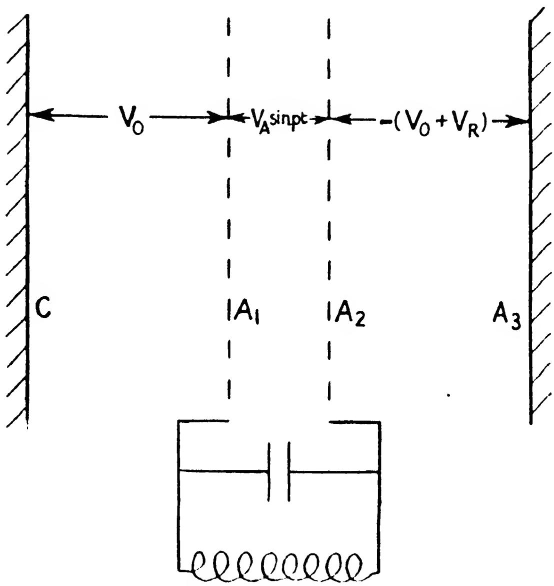


Fig. 24. A schematic representation of a reflex oscillator

direction. During their stay in the retarding field, the electrons form bunches by reason of their varying initial velocities, and these bunches, traversing the gap A_2A_1, induce currents in the circuit attached thereto. If this current has the correct phase relationship to the voltage across A_1A_2, and if the power fed into the circuit is greater than that dissipated by the resistance, then the oscillations are self-maintained. This is a qualitative description of the mechanism of the reflex oscillator.

As in the discussion of the klystron oscillator, let the average electron emerge from the gap at $t = 0$, where $V = - V_A \sin pt$. Then this electron becomes the centre of a bunch which returns to the gap at time $t = \tau$. For oscillation to occur, this time τ must be such that the voltage across the gap is $V = V_A$, for it is at this epoch that the maximum energy is taken from the bunch. Therefore $p\tau = 2\pi(n - \frac{1}{4})$. The average electron emerges into the retarding field with velocity $v_0 = \sqrt{2V_0e/m}$. In the space A_2A_3 the electronic deceleration is α where $m\alpha = e(V_0 + V_R)/d$ and d is the distance A_2A_3.

Therefore $$\tau = \frac{2v_0}{\alpha} = \frac{2\sqrt{2V_0e/m}d}{(V_0 + V_R)e/m}$$

and the phasing condition becomes

$$V_0 + V_R = \sqrt{\frac{2V_0}{e/m}}\frac{pd}{\pi(n - \frac{1}{4})} \quad . \quad . \quad . \quad . \quad (2.23)$$

In general the retarding field is not constant and the above theory does not apply in this simple form. But it serves to indicate that, as in the case of the two-circuit klystron, the reflex variety oscillates in the vicinity of discrete values V_0 and V_R. An analysis of the beam current made in the manner described in Section 2.4.1.1 shows that the current has the same form as in the case of the two-circuit klystron, so that equation (2.17) holds equally for the reflex case: for the time of flight of an electron which enters the reflecting field at time $t = t_1$ is τ, where

$$\tau_0 = \frac{2dv_0}{(V_0 + V_R)e/m}\left[1 + \frac{V_A}{2V_0}\sin pt_1\right], \text{ if } \frac{V_A}{V_0} \text{ is small.}$$

The time of flight of the electron entering at time $t = 0$ is T_0

where $$T_0 = \frac{2dv_0}{(V_0 + V_R)e/m}$$

Therefore $$\tau = t_2 - t_1 = T_0\left(1 + \frac{V_A}{2V_0}\sin pt_1\right)$$

Compare this with equation (2.15).

Therefore $$\frac{I_2}{I_1} = \frac{dt_1}{dt_2} = \frac{1}{1 + \frac{pT_0V_A}{2V_0}\cos pt_1}$$

and, as in the case of the two-resonator klystron, this leads directly to the value of I_2, the alternating current at the catcher. The fundamental component of I is I_1 where

$$I_1 = 2I_0J_1\left(\frac{\theta V_A}{2V_0}\right)\cos(pt - \theta) \quad . \quad . \quad (2.17\text{A})$$

and $$\theta = pT_0 = \frac{2pdv_0}{(V_0 + V_R)e/m}$$

In the next chapter the reflex klystron oscillator will be considered in more detail, for it is an important element in micro-wave communication systems.

2.4.2. *Velocity Modulation Valves.* In this section the valves described will be considered only as an assembly of electrodes, producing and controlling an electron beam. The circuits attached to the valves, although these often form an integral part of the structure, will be examined in the following chapters. This somewhat artificial procedure is adopted to emphasize that the circuits obey the usual laws and may be altered without changing the valve electrodes to satisfy different requirements.

2.4.2.1. *The Klystron.* The two-gap klystron is made in many forms. Using the copper disc-seal technique, a typical arrangement of electrodes is shown in Figure 25 which is a section of a valve having cylindrical symmetry about the vertical axis. At one end of the glass envelope G, next to the pinch P, is the electron gun which produces the beam through the tube. This gun in its simplest form consists of a disc or button cathode of the oxide-coated variety, surrounded by a focusing cylinder S, and facing the accelerating electrode E_1. The focusing cylinder is generally maintained at a negative potential with regard to the cathode, although in cases where the modulation of the output is performed entirely by varying the anode potential at which E_1, E_2 and E_3 are maintained, the cylinder S is at cathode potential. The electrode E_1 in addition to being the anode of the

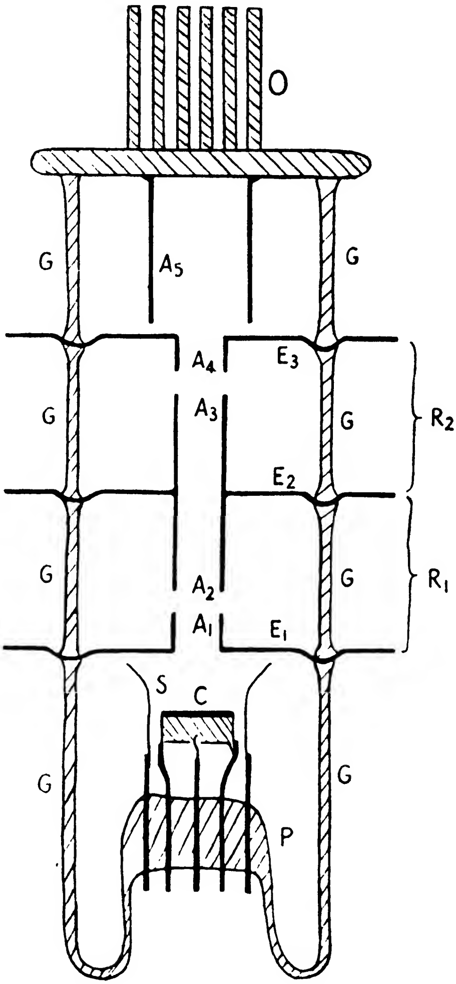


Fig. 25. An axial section of a typical disc-seal klystron valve

gun is one of the discs forming the buncher circuit which is connected between E_1 and E_2. This circuit in common with the catcher circuit is usually some form of modified concentric line, having cylindrical symmetry about the axis of the tube.

Figure 25 should be compared with Figure 22 to understand the corresponding structures. The buncher gap A_1A_2 is designed so that the coupling coefficient β is of the correct magnitude, and the modulated beam proceeds along the drift tube A_2A_3 to the catcher gap A_3A_4. In some klystrons the electrons are collected on the electrode E_3, but with copper discs, to avoid undue temperature rise in the catcher circuit, an additional collecting electrode A_5 is added with cooling vanes O attached.

Figure 25 is to no particular scale, but represents the form taken by a typical copper disc-seal valve. The collecting electrode A_5 is brazed to a nickel iron disc, forming the end-plate of the tube. Klystrons of this type were made in Admiralty Signal Establishment during World War II for use by all three Services.

The problems of klystron design are manifold, but one of the major preoccupations of the engineer is to obtain a suitable beam of electrons. As will be seen in the next chapter, when amplifiers and oscillators are discussed, the greater the current density which can be obtained without undue loss to the electrodes E_1, E_2 and E_3, the more efficient the klystron becomes.

Another point worth noting about Figure 25 is that the glass wall of the tube is in the buncher and catcher circuits. The losses in the glass become the limiting factor at frequencies in the neighbourhood of 5000 to 6000 Mc/s, when a new design must be employed.

Klystrons made in U.S.A. have employed an entirely different technique, where the buncher and catcher circuits themselves form the envelope of the valve. Such tubes will work at higher frequencies than those described above, for there is no insulator loss in the circuits.

2.4.2.2. *The Reflex Klystron.* Figure 26 is a diagram of a typical copper disc-seal reflex klystron without its circuit. Again there is cylindrical symmetry about the tube axis. The electron gun consists as before of a cathode C, a focusing cylinder S and an accelerator E_1, the latter being one electrode of the circuit. The gap A_1A_2 is arranged so that the electrons move directly from the bunching field into the retarding field without the intervention of a force-free space. If a force-free space is present, the bunching mechanism is similar to that in the two-circuit klystron, and opposes the mechanism for the formation of bunches in a retarding field. So it is bad design in a reflex oscillator to have the gap symmetrically placed with regard to E_1 and E_2. It should, on the contrary, be as close to E_2

as possible. A_3 is the reflector which should be formed so that it constrains the beam to re-enter the gap A_2A_1 with the minimum loss to E_2. It should at the same time cause the beam to be collected on the cylindrical wall below the gap, for, if the electrons return to the vicinity of the cathode, multiple transits may result, and these are the cause of certain unfavourable phenomena in the oscillator. In particular, a valve in which there are multiple transits has a high noise content, and is unsuitable for use as a receiver oscillator.

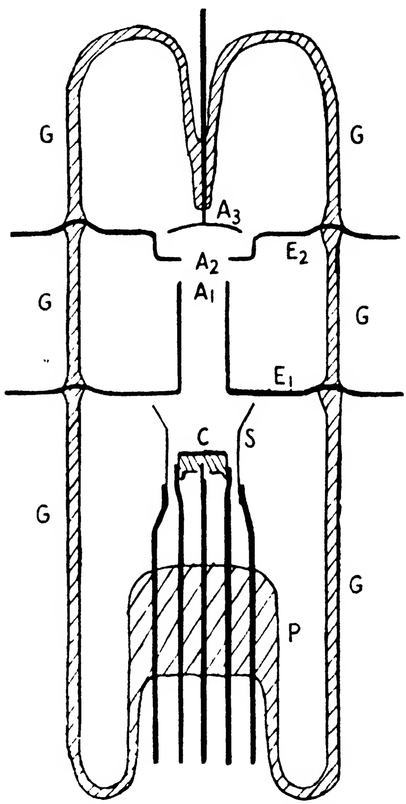


Fig. 26. An axial section of a disc-seal reflex valve

As in the case of all disc-seal valves, the upper frequency at which the device will operate is determined by glass losses. American practice is again to eliminate the glass walls and to have the complete oscillatory circuit inside the envelope, or part of the envelope. The British practice at higher frequencies is to arrange the glass at a part of the circuit where there is the minimum potential difference across it and therefore the minimum loss. This is the so-called "nodal seal" technique.

2.5. **The Magnetron Oscillator.** Magnetrons have been used in U.H.F. radio communication, but it is usually because a more satisfactory valve does not exist. The magnetron is essentially an oscillator, consisting of a single circuit coupled to an electron beam. In this respect it corresponds to the reflex klystron. But self-oscillators are not much used in communication transmitters, and as a receiver oscillator the magnetron is "noisy." For that reason the future use of the magnetron may be confined to radar transmitters. Multiplex pulse communication systems can employ the magnetron oscillator where frequency control need only be very rudimentary, and this will be discussed later in the appropriate section. But in view of the limitations of space, no detailed examination of the magnetron will be made in this chapter.

2.6. The Travelling Wave Tube. One of the greatest difficulties encountered in the use of klystron amplifiers is that the working Q of each circuit is necessarily large. This is most marked in the multi-circuit klystron amplifier (see Chapter III), but it is true of all. This high effective Q means that the klystron amplifier is a comparatively narrow-band device, requiring accurate tuning of all its circuits. This is a difficult and expensive procedure in radio communication, particularly where the transmitter has to cover a wide frequency band. It may, indeed, make the device useless for some

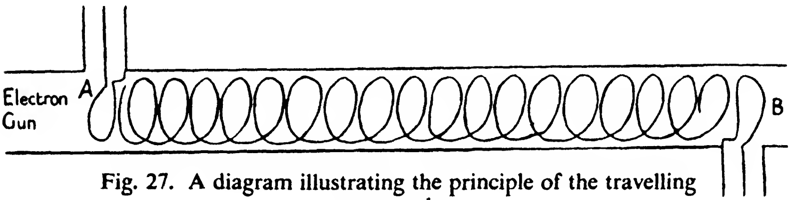


Fig. 27. A diagram illustrating the principle of the travelling wave tube

applications. In 1942 Kompfner[9] enunciated the principle of the travelling wave amplifier which is *a priori* broad-band. The elementary theory of the device is as follows—

A signal is inserted at A (Figure 27) into a concentric line circuit in which the inner conductor is wound in toroidal form. An electron beam, magnetically focused, traverses the toroid axially. The signal voltage modulates the beam and the electrons move along the tube at a speed dictated by the accelerating voltage in the electron gun. Owing to the toroidal form of the inner conductor of the line, the signal is propagated along AB with a velocity of the same order of magnitude as the electrons. The interaction of the signal and the beam continues to the end of the tube at B, where the remaining signal is extracted from the tube. Under certain conditions of voltage and frequency, the U.H.F. power taken from the kinetic energy of the electrons due to the accelerating voltage can be considerable, and the tube then operates as an amplifier. To avoid regeneration it is necessary to insert attenuation in the line. This can be done without interfering with the mechanism of amplification.

Tubes of this nature are now in the experimental stage and are being developed in America, France and Britain, but it is still too early to estimate their value. The length of a travelling wave amplifier at 3000 Mc/s is considerable, and its associated magnetic field, developed by means of a toroid, is inconvenient. These difficulties may be overcome in the future. Travelling wave amplifiers can be transformed into oscillators by the usual methods, and these have possibilities at the highest frequencies. In common with klystrons they appear to be noisy as signal amplifiers or frequency changers, and so their immediate application is to transmitters.

REFERENCES

(1) *Phil. Mag.*, March 1928, Feb. 1931.
(2) "Electron Inertia Effects," Cambridge Physical Tracts, 1941.
(3) *Hochf. u. Electroa.*, 41, May 1933 and 46, Nov. 1935.
(4) Llewellyn, *loc. cit.*
(5) *Nature*, 157, p. 841 (1946).
(6) *Nature*, 159, p. 640 and p. 677 (1947).
(7) Bakker and de Vries, *Physica* 1, p. 1045 (1934); Begovich, *J. App. Phys.* 20, p. 457 (1949).
(8) *J. App. Phys.* 10, p. 501 (1939).
(9) *Wireless Engineer*, 24, p. 255 (1947).

Chapter III

POWER AMPLIFIERS, FREQUENCY MULTIPLIERS AND OSCILLATORS

In this chapter consideration will be given to the units which constitute the U.H.F. stages of a communication transmitter. Two widely different techniques have been developed. In one type of communication system the self-oscillator operates at a comparatively low frequency and the U.H.F. stages consist entirely of amplifiers and frequency multipliers. A common example of this type is the transmitter the frequency of which is controlled by means of one or more piezo-electric oscillators working at frequencies in the neighbourhood of a few megacycles. In the other type of system the self-oscillator is itself in the U.H.F. region and incorporates its own frequency control in the form of a high Q resonator or line circuit. Methods of frequency control will be discussed in detail in Chapter V, and in view of the fact that all transmitters fall into one of the two categories mentioned above, it is possible to confine attention in the present chapter to units which find an immediate application in one or other of the systems. For this reason it is not necessary to discuss certain U.H.F. oscillators which have important applications outside the communication field, and this is the only mention which will be made of two-circuit klystron oscillators, magnetrons and triode high-power oscillators.

3.1. **Triodes.** The frequency range in which triodes may be used has been enormously extended by the development of the disc-seal valve, and transmitting units incorporating triodes can now be employed with profit at frequencies up to 2500 Mc/s. Most present-day triodes have begun to lose their efficiency to a marked degree at this frequency but there is no reason to doubt that the upper limit will be pushed to a much higher value in the future.

3.1.1. *Power Amplifiers.* At U.H.F. power amplifiers are always operated in the common-grid condition described in Chapter II. The theory of operation, so far as the anode-grid circuit is concerned, is similar to the conventional power amplifier theory, except that there is an additional transit-time damping across the circuit. This is usually small at frequencies below 2000 Mc/s, and does not constitute the limiting factor. Hence anode efficiencies of the same order of magnitude as those obtained at lower frequencies can be achieved in the metric wave-length region. At $\lambda = 50$ cm. it is quite possible to use a triode at an anode efficiency of 60 to 70% in the

C.W. condition. A typical disc-seal triode used in a concentric line circuit similar to that shown in Figure 16 can be adjusted to operate at an efficiency of 75% at frequencies of the order of 500 Mc/s. Between 1000 and 2000 Mc/s the efficiency diminishes rapidly, the value depending upon the size of the valve. Power amplifier "tank" circuits may be balanced or unbalanced and the circuit illustrated in Figure 15 is suitable for use at frequencies up to 600 Mc/s.

Common-grid amplifiers at U.H.F. suffer from two disabilities. The first, the inevitable negative feed-back of the common-grid circuit, would not be serious by itself, for the increase in driving power which is its result is partially compensated by the appearance of that driving power in the output. But the second disability, transit-time damping, or "total emission damping," mentioned in Chapter II, is a much more serious disadvantage, and taken along with the negative feed-back, severely limits the power gain which is attainable. It is not yet possible to give any quantitative analysis of U.H.F. power amplifiers, since the total emission damping is not sufficiently understood, but the experimental facts are well-known and should be found useful. The stage gain of a power amplifier of the type under discussion is not usually greater than 10 db, and the maximum gain often occurs at less than full dissipation. The design of disc-seal triodes is indeed to some extent a matter of balancing the cathode emission against the anode dissipation. As the frequency increases, the power gain diminishes at first slowly and then rapidly. It is only in exceptional valves that the gain is much above unity at 3000 Mc/s.

British disc-seal valves are made in several sizes and at the moment of writing have been mainly developed by the General Electric Company at their Research Laboratories in Wembley. The power-amplifiers in the present series range from the CV273, which has an anode dissipation of 6 watts and a gain greater than unity at 3000 Mc/s, to a valve the anode dissipation of which is 2 kW and which will give 1·5 kW at 200 Mc/s. This is a useful range of valves for communication purposes, and shows that the future designer need not fear lack of power for almost any purpose.

One of the most important parts of a U.H.F. power-amplifier is the output coupling arrangement. There are two possibilities, as in the more conventional circuits at lower frequencies: either electrostatic or magnetic coupling may be used. But the output is usually fed into a concentric line, and the characteristic impedance of the line is usually between 50 and 100 ohms. If the amplifier is single-ended, the problem is to match the load into the circuit and that is done by well-known methods of which double-stub matching is possibly the best. If the amplifier is balanced, the unbalanced output requires careful design to avoid losses due to standing waves on the outer conductor and resistive losses due to unbalancing the tank circuit. Usually, it is more satisfactory in this case to employ electrostatic coupling, balancing the reflected impedance of the output circuit on one side of the line by a corresponding capacitance on the other side. Using this method more than 90% of the power can be coupled into the load, and the balance is not sufficiently frequency sensitive to cause difficulties unless the frequency-band to be covered by the amplifier is of the order of an octave. This is very rare. Care must be taken to see that the balancing capacitative reactance is not lossy, for the out-of-phase current flowing in it can be very high indeed.

3.1.2. *Frequency Multipliers.* Common-grid frequency multipliers are very inefficient, but have the useful property that they can be

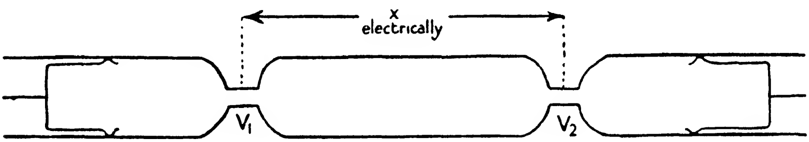


Fig. 28. Illustrating how two valves may be used to excite a single cavity resonator

employed at higher "tank" circuit frequencies than power amplifiers. This is because the inefficiency of the amplifier is largely due to losses in the cathode-grid circuit, and where the unit is a multiplier the cathode-grid losses correspond to a sub-multiple of the tank circuit frequency. It is, however, impossible to extend the frequency range very far by this effect, for the glass losses in the grid-anode space become important at frequencies above 3000 Mc/s. Nevertheless, the author has used triodes as treblers to give a useful output for frequency measurement (a few milliwatts) at 5000 Mc/s, the circuit being similar to that of Figure 8. He has also communicated over a distance of 20 miles at a frequency of 3000 Mc/s, using a transmitter the final stage of which was a triode trebler of anode dissipation 6 watts. So it is not reasonable to dismiss the triode even at the highest frequencies in the U.H.F. range. Certain remarkable circuits can be made for multipliers and amplifiers, using triodes across rectangular guides and operating the $TE_{1,0}$ mode of oscillation. Figure 28 is an illustration of such a circuit. Imagine a guide into which two triodes V_1, V_2 have been inserted as in Figure 8, the electrical distance between their centres being x cm. Let each be driven by a concentric line cathode-grid circuit. Then consider (i) an amplifier. If $x = \lambda_g$, the wave-length in the guide corresponding to the frequency of excitation, then the guide is the common tank circuit of two valves in parallel. One can imagine a continuous ring of guide containing several triodes at the appropriate points. Now consider (ii) a doubler. Let the distance $x = \lambda_g$, the wave-length corresponding to the frequency $2f$ where f is the frequency of excitation. Then if the cathode-grid circuits are driven with a phase difference of 180° (as they would be, for example, by means of a shielded pair, one going to each cathode) the circuit corresponds to a "push-push doubler."

Many other fascinating possibilities may be imagined, and the system has considerable value so long as the glass losses are not excessive. It is, of course, only applicable to "spot frequency" working.

It is not essential to operate disc-seal triodes in common-grid circuits, when they are employed as multipliers in the lower half of the U.H.F. spectrum. The regeneration which occurs in a common cathode amplifier at these frequencies does not affect a multiplier. But experiment has shown that when the circuits are properly designed there is little difference in efficiency and stage gain between the two types of operation. The all-important factor—grid damping—applies equally to common cathode and common grid. Another interesting point about U.H.F. multipliers should be mentioned. As the frequency increases, it becomes physically impossible to realize a quarter-wave line which includes the valve capacitances. The circuit would be totally inside the glass envelope or so small that it could not be fitted. If the fundamental wave-length is λ and the unit is a frequency doubler, then the effective length of the cathode line would be $\lambda/4$ and the effective length of the anode line would be $\lambda/8$. If the latter is physically unrealizable, it is possible to work at the 3/4 wave-length mode, the length of the anode line being effectively $3\lambda/8$. This is difficult to arrange at lower frequencies, where the Q of the tank circuit may be such that the $\lambda/4$ mode (amplification) and $3\lambda/8$ mode (doubling) run together. Where the Q is due to a well-made concentric line, there is usually no difficulty in separating the two actions. Similar tricks can be played with a trebler, but in this case the 5/4 wave-length mode must be used, for the 3/4 mode would be amplification of the fundamental. At frequencies where a wave-guide can be used for the anode circuit, the $3\lambda/4$ mode can be operated in trebling, since the dimensions of the guide may be chosen to attenuate the fundamental in an extremely short distance.

3.1.3. *Oscillators.* Some general observations on the use of U.H.F. oscillators in radio communication are necessary. As at lower frequencies, the modulated self-oscillator is not satisfactory, and for the same reasons. In all communication systems the engineer strives to employ the oscillator only as a reference device to determine the position of his carrier on the frequency scale. Where amplitude modulation is used, if a U.H.F. oscillator appears at all, it is a low-power device followed by a buffer amplifier and one or more power amplifiers. Where phase or frequency modulation is used, the frequency determining oscillator may be modulated directly, but if it is feasible, the power is subsequently raised to a higher level by amplifiers. Pulse modulation, which inevitably requires a wide frequency-band and therefore does not involve such precise frequency determination, may be applied to the oscillator, and in this case the oscillator may be of such high power that no additional amplifier is used. Magnetrons, for example, are useful pulsed oscillators for this type of communication, and indeed, in the high regions of the spectrum satisfactory pulse amplifiers do not exist. But the aim of the communication engineer must always be to separate the oscillator from the modulator, and the above-noted examples of present techniques represent what can be done rather than what should be done. On the receiver side U.H.F. oscillators of low power are used in conjunction with rectifiers to transfer the signal to a carrier in a more convenient region of the radio spectrum where the necessary amplification may be attained. In this case also, the frequency stability of the self-oscillator is a matter of great importance; and where space and power permit, it is better technique to use as a local oscillation a multiple of a frequency generated in a self-oscillator operating at a much longer wave-length. In some cases, however, a very wide band of frequencies must be covered without gaps by the receiver, and then a U.H.F. oscillator is almost a necessity. This may be a triode or a reflex klystron, according to the range to be covered. The triode oscillator will now be discussed.

Common-grid amplifier stages make satisfactory oscillators, when a controlled amount of positive feed-back is introduced between the anode and the cathode circuits. In employing concentric lines, the feed-back generally takes the form of the capacitative action of a probe extending through the common cylindrical wall of the two circuits. The position and length of this probe may be adjusted to give a uniform performance over a certain frequency range. Where the range is large, some automatic means of varying the "back" coupling can be introduced. In some disc-seal triodes a probe is inserted in the electrode structure, connecting the anode and cathode spaces. This has the advantage of eliminating a variable factor from the external circuit, but it is an inflexible design which may limit the performance of the oscillator under certain circumstances. In general the engineer will probably prefer to use an amplifier valve and to adjust the feed-back to suit the remainder of his circuit. Butterfly circuits make excellent oscillator elements, and only one need be employed (between the anode and grid), the power being fed back to the cathode by means of a small capacitance. Such butterfly oscillators, if not too heavily loaded, will operate over a very wide range of frequencies, but if the load is optimised they have a tendency to go out of oscillation at certain spots in the band. These "blind spots" are generally due to unwanted resonances of components forming the unit, but it is difficult to eliminate them completely even with the most careful design.

The short-term frequency stability of a lightly loaded U.H.F. triode oscillator can be relatively high, particularly if a good line circuit is connected across grid-anode. The "long lines" stabilization effect, commonly used at lower frequencies, is not, however, applicable to grounded-grid oscillators, for the damping due to the series feed added to whatever transit-time damping may exist, makes the input circuit a relatively poor frequency determining element. Nor is it possible to apply automatic frequency control to such oscillators, for a reactor valve is only a theoretical conception at U.H.F. As a consequence, triodes are only used as oscillators in comparatively wide-band receivers. Their chief defect at the present time is their long-term frequency drift with variation of temperature. Disc-seal valves are serious offenders in this respect, for a change in the anode dissipation, leading to a change of temperature within the glass envelope, can cause an appreciable change in the inter-electrode capacitances. This causes frequency change which cannot be easily counteracted by the external circuit.

3.2. **Velocity Modulated Valves.** These can again be classified as oscillators, amplifiers and frequency multipliers, and all three exist as practical transmitting devices. On the oscillator side, only the reflex klystron will be discussed for reasons already mentioned. Two- and three-circuit klystrons will be considered both as amplifiers and as multipliers.

3.2.1. *Reflex Klystrons.* The reflection klystron has already been discussed with special reference to its electron beam in Chapter II. It forms a most useful communication component at centimetric wave-lengths, both as a receiver local oscillator and as a part of the frequency determining unit of a transmitter. The circuit generally employed in conjunction with a reflex klystron, and as an integral part of the valve structure, is a modification of the concentric line. Figure 29 shows a typical example.

The circuit (resonator) which has cylindrical symmetry is shown in axial section by ABCDEFGHIJ, the part corresponding to the inner conductor being the cylinder ABJC and the short cylinder GF. These two are separated by a gap, corresponding to A_1A_2 of

Figure 24. The electron gun is represented by the section of the cylindrical cathode N facing the grid JC and surrounded by the focusing cylinder OO′. The reflecting electrode is R. The grids, stretched across the cylindrical apertures allow the majority of the

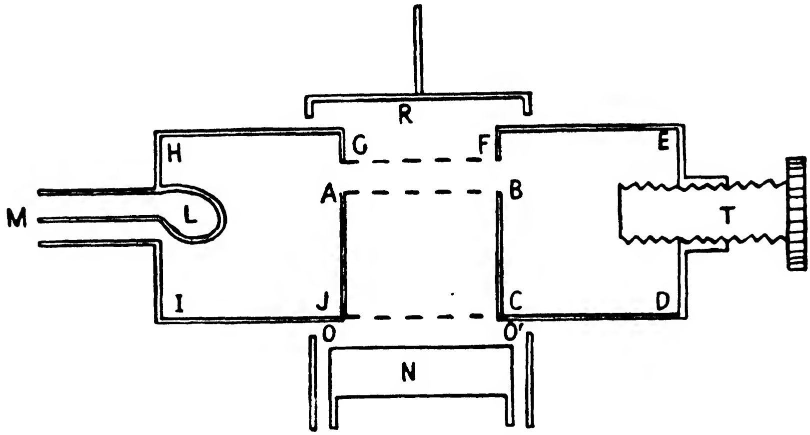
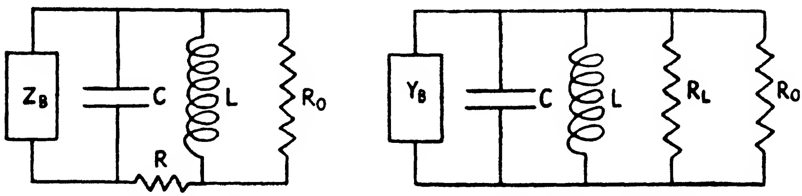


Fig. 29. An axial section of a reflex oscillator, showing the resonator with its tuning plunger and output coupling loop

electrons to pass, but preserve the strong electric field which produces a high coupling coefficient β. The concentric resonator may be tuned by means of a plug T which, moved radially towards the axis, initially diminishes the effective lumped inductance of the circuit, and therefore increases its natural frequency, but finally, if



Figs. 30 and 31. Equivalent circuits of the reflex oscillator

moved very close to the gap, increases the effective lumped capacitance and therefore decreases the frequency. The U.H.F. power from the oscillator is led away through the loop L, coupled to the field in the resonator, by the concentric line M.

There are, of course, many different practical forms taken by reflex klystrons, but the principles of operation are always the same. It is with these principles that we are now concerned.

Let the reflex oscillator be represented by an impedance Z_B shunted by the circuit and the load. Then the equivalent circuit is as shown in Figure 30, where C, R, L are the equivalent lumped constants of the resonator and R_0 is the load resistance. To make all the quantities additive, write $Y_B = 1/Z_B$ and insert the shunt resistance R_L in place of the series resistance R (Figure 31). For oscillation to occur, the real part of Y_B must be negative and greater than the conductance $\frac{1}{R_L} + \frac{1}{R_0}$. The voltage across the gap $V_A \sin pt$ gives rise to the current $I_1 = 2\beta I_0 J_1(q) \sin(pt + \phi)$ (equation (2.17A), paragraph 2.4.1.8) and therefore the beam conductance $1/R_B$ is given by

$$R_B = \frac{V_A}{2\beta I_0 J_1(q) \cos\phi}$$

Omit β for simplicity. Then for oscillation to occur,

$$\frac{2I_0J_1(q)\cos\phi}{V_A} > \left(\frac{1}{R_L} + \frac{1}{R_0}\right) > \frac{R_0 + R_L}{R_0R_L} > \frac{1}{R_E}$$

Now $q = \theta V_A/2V_0$, therefore $V_A = 2V_0q/\theta$

Therefore when oscillation just begins,

$$\frac{q}{J_1(q)} = \frac{\theta I_0 R_E \cos\phi}{V_0} \quad . \quad . \quad . \quad . \quad (3.1)$$

But in this régime, V_A/V_0 is very small and therefore q is small and $J_1(q) \doteq \frac{1}{2}q$. Also $\cos\phi = 1$ for exact phasing. Therefore equation (3.1) becomes

$$I_0 = \frac{2V_0}{\theta R_E} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.2)$$

which gives the mode of variation of I_0, the starting current, with V_0, θ and R_E. But $\theta = 2\pi(n - \frac{1}{4})$ (paragraph 2.4.1.8)

whence

$$I_0 = \frac{V_0}{\pi R_E(n - \frac{1}{4})} \quad . \quad . \quad . \quad . \quad (3.3)$$

Power Output of the Reflex Oscillator. Consider the power P delivered to the resistance R_E when the beam current has a fundamental component $2I_0J_1(q)\sin(pt + \phi)$ and is associated with a gap voltage V_A. Let $\phi = 0$, i.e. let the reflector voltage be adjusted for perfect phasing, and let $\beta = 1$ as before.

Then $\quad P = \frac{1}{2}V_A \,.\, 2I_0J_1(q) = V_AI_0J_1(q)$

But $\quad V_A = 2V_0q/\theta$

Therefore $\quad P = \frac{2V_0I_0}{\theta} qJ_1(q) \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.4)$

As in any oscillator, the power P will adjust itself so that the (negative) admittance of the beam current is equal to the (positive) admittance of the circuit, and if $\phi = 0$, this gives

$$\frac{I_0 J_1(q)\theta}{V_0 q} = \frac{1}{R_E} \text{ or } \frac{J_1(q)}{q} = \frac{V_0}{I_0 R_E \theta} \quad . \quad . \quad . \quad (3.5)$$

Formally, q can be eliminated between equations (3.4) and (3.5), giving P as a function of V_0, I_0, R_E and θ. In practice this can be done for any given values of the constants by using a table of $J_1(q)$. For example, consider the case where $V_0 = 300$ volts, $I_0 = 10$ mA and $R_E = 10{,}000$ ohms. Then, immediately,

$$\frac{J_1(q)}{q} = \frac{3}{2\pi(n - \frac{1}{4})} \quad . \quad . \quad . \quad . \quad (3.6)$$

Table V below gives details of the calculation where the object is to determine the output power in the various modes obtained by varying the reflector voltage only. Row 2 is evaluated from equation (3.6) and row 3 from a graph of $J_1(q)/q$ (Figure 32). Row 4 is found from a graph of $qJ_1(q)$ (Figure 33) and row 5 from equation (3.4).

TABLE V

1.	n	1	2	3	4	5	6
2.	$J_1(q)/q$	0·637	0·273	0·174	0·127	0·101	0·083
3.	q	—	2·080	2·645	2·920	3·085	3·195
4.	$qJ_1(q)$	—	1·186	1·212	1·070	0·948	0·840
5.	P(mW)	—	647	421	273	190	140

Three points should be noted. First, the beam current is too small to sustain an oscillation in the mode $n = 1$. Second, the values in row 5 are *relatively* fairly correct, but the absolute scale is not realistic. This is due to a number of causes of which two are outstanding. First, the beam current has been given an effective value of 10 mA. In any practical low-voltage reflex klystron this would imply a total current drawn from the cathode of the order of 50 – 100 mA, for the transparency of each grid is usually of the order of 80–90%. Second, a β-factor of unity has been assumed, and since P is proportional to β^2, a correction on this account is necessary. In practice, overall efficiencies of the order 2–5% are obtained from reflex oscillators. Finally, nothing has been said of the losses in the circuit. It has been assumed throughout that the oscillator is heavily loaded. This is seldom possible in practice.

In the general case, of which the above calculation is a particular example, account must be taken of the reactive as well as of the resistive action, and the current must be written

$$I = 2I_0 J_1(q) \sin(pt + \phi) \quad . \quad . \quad (3.7)$$

where the voltage across the gap is

$$V = V_A \sin pt \quad . \quad . \quad . \quad . \quad (3.8)$$

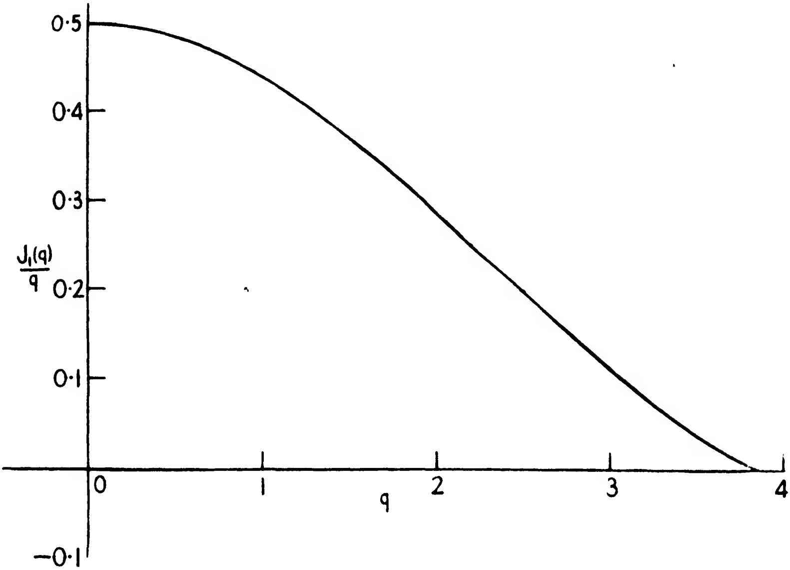


Fig. 32

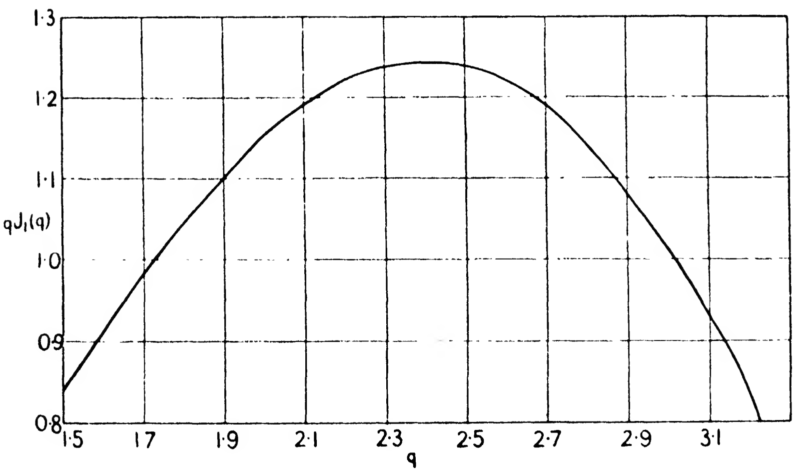


Fig. 33

Then
$$P = V_A I_0 J_1(q) \cos\phi$$
$$= \frac{2V_0 I_0 q J_1(q) \cos\phi}{\theta} \quad . \quad . \quad . \quad (3.9)$$

when
$$\frac{I_0 J_1(q)\theta \cos\phi}{V_0 q} = \frac{1}{R_E} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.10)$$

and
$$\theta = \frac{2dv_0 p}{(V_0 + V_R)e/m} = 2\pi\left(n - \tfrac{1}{4} + \frac{\phi}{2\pi}\right) \quad . \quad . \quad . \quad (3.11)$$

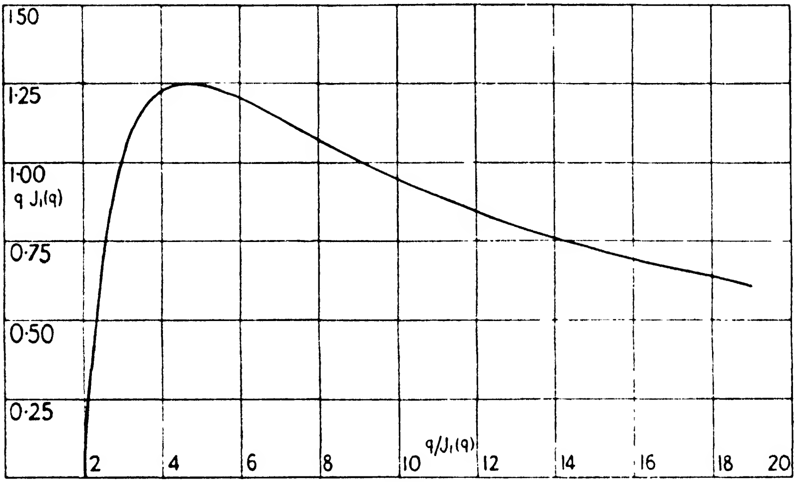


Fig. 34

At the same time
$$\frac{I_0 J_1(q)\theta \sin\phi}{V_0 q} = \frac{\sin\phi}{R_E \cos\phi} = -\frac{1}{X_E} \quad . \quad . \quad (3.12)$$

where X_E is the beam reactance.

To obtain P from the value of R_E, or, in fact to perform almost any operation on these equations entails finding $qJ_1(q)$ in terms of the corresponding value of $q/J_1(q)$. It is therefore worth while to construct a graph of the one function plotted against the other. This is shown in Figure 34.

Consider now the operation of the reflex oscillator in the vicinity of a value of θ such that $\phi = 0$, i.e. consider the variation in power and frequency to be expected when the reflector voltage (for example) is varied about a value which gives a maximum output. This is an important matter as will be seen in Chapter IV when frequency control is under examination.

Once more, let $V_0 = 300$ volts, $I_0 = 10$ mA, $R_E = 5000$ ohms and let θ vary about the $n = 3$ mode, its values ranging from 5π to 6π.

Then, using equations (3.9) and (3.10), we find that the valve oscillates for values of ϕ between $-41°$ and $+44°$ and that the power output varies as shown in Figure 35 (*a*).

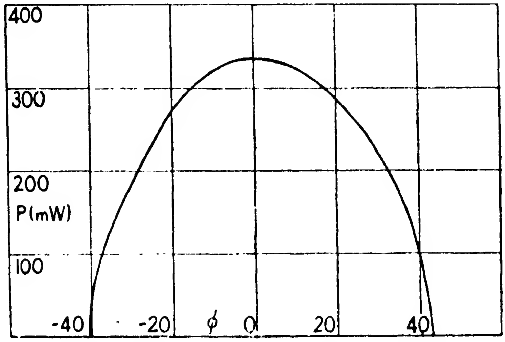


Fig. 35 (*a*). The variation (*a*) of the power output and (*b*) of the frequency of oscillation with the phase angle of the returning electron bunch

Using equation (3.12) and equating susceptances in the circuit of Figure 31,

$$-\frac{1}{pL} + pC - \frac{\tan\phi}{R_E} = 0$$

Let $\tan\phi = 0$ correspond to $p = p_0 = 2\pi f_0$, where f is the frequency of oscillation. Then

$$\left(\frac{p}{p_0}\right)^2 - 1 = \frac{\tan\phi}{Q_E}$$

for the loaded Q is $$Q_E = \frac{R_E}{pL}$$

Let $p = p_0 + \delta p$ and let δp be small. Then to the correct order of approximation,

$$\left(\frac{p}{p_0}\right)^2 - 1 = \frac{2\delta p}{p_0} = \frac{2\delta f}{f_0}$$

and $$\delta f = \frac{f_0}{2Q_E}\tan\phi \qquad (3.13)$$

In the example under consideration ϕ varies from $-41°$ to $+44°$. Assume $Q_E = 300$. Then the variation of $\delta f/f_0$ is shown in Figure 35 (*b*).

Hence, by varying the potential of the reflector, so that the phase of the returning electrons gives rise to a reactive component of impedance, the frequency of oscillation may be varied in an almost linear manner. As might be expected, the variation of frequency for a given ϕ is inversely proportional to the loaded Q of the circuit. This property of the reflex oscillator has important consequences. It allows it to be used (*a*) in conjunction with a resonator as a

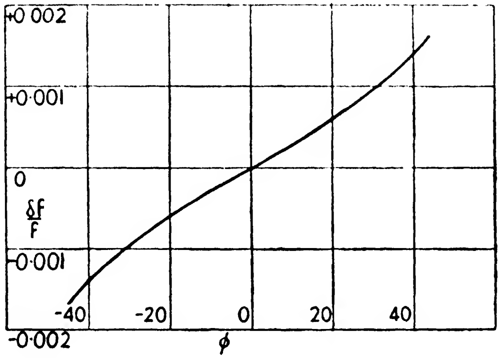


Fig. 35 (*b*)

frequency determining element of a transmitter or receiver and (*b*) in conjunction with a source of audio-frequency voltage as a frequency modulator.

3.2.2. *Klystron Power Amplifiers.* The velocity modulation effects in a klystron amplifier have already been discussed in Chapter II,

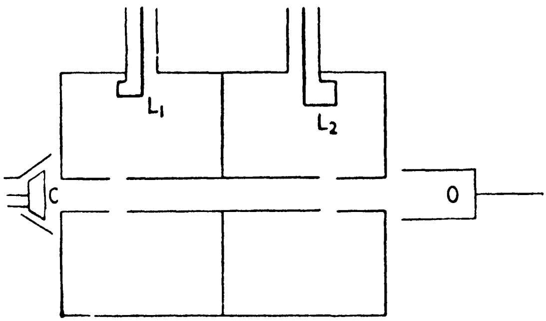


Fig. 36. An axial section of a klystron amplifier

and Figure 25 is a diagram of the electrode arrangement in a two-circuit valve. The circuits generally consist of resonators of the loaded concentric line type. Figure 36 is an axial section of two such circuits as they would be arranged in a simple power amplifier.

The input to the "buncher" resonator is generally from a coaxial cable by means of a loop (L_1), and the output power is led away by a similar loop (L_2). The greater part of the beam should be collected on the isolated electrode O, so that the temperature of the resonators may be maintained at a constant level independent of the operation of the valve. In practice this is only partially achieved, and in certain types of klystron the resonators may be water-cooled.

A typical two-resonator amplifier might operate with a beam voltage (V_0 of paragraph 2.4.1.1) of 2500 volts and a current of 50 mA, at an anode efficiency of 12%, giving an output of 15 watts. At this level the power gain would be of the order of 10 db. At

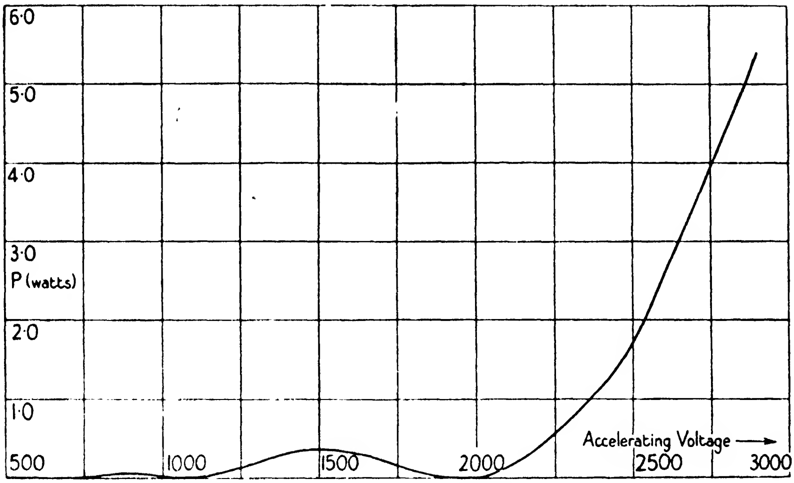


Fig. 37. Variation of the power output of a klystron with the accelerating voltage

lower inputs the gain might be higher, as may be seen theoretically by examining the mode of variation of $J_1^2(a)$ with a. Of more interest to the communication engineer, however, is the behaviour of the amplifier for a given radio-frequency input when the accelerating voltage, V_0, is varied.

From Chapter II the power developed in a load resistance R_0 which we shall assume to be small compared with the shunt resistance of the circuit is given by P, where

$$P = 2\beta^2 I_0^2 J_1^2(a) R_0 \quad . \quad . \quad . \quad (3.14)$$

and

$$a = \frac{pxv_1}{v_0^2} = \theta\frac{v_1}{v_0} = \theta\sqrt{\frac{V_A}{V_0}}$$

If, therefore, it is required to examine the mode of variation of P with V_0, equation (3.14) may be rewritten

$$P = KV_0^3 J_1^2\left(\frac{b}{V_0}\right) \quad . \quad . \quad . \quad (3.15)$$

remembering that $I_0 = kV_0^{3/2}$. The factors K and *b* are arbitrary constants.

Figure 37 exhibits P as a function of V_0 according to equation (3.15), the values of P being adjusted on an absolute scale to correspond roughly to a klystron of this type working with the correct radio-frequency input.

The most interesting feature of Figure 37 is that there is a considerable region on the scale where the output varies almost linearly with the voltage, and although in practice variations from this simple relationship are to be expected, the general shape of the curve is well predicted by theory.

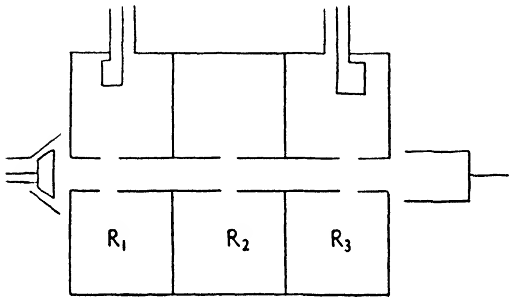


Fig. 38. An axial section of a three-resonator klystron amplifier

It will be interesting in Chapter V to return to the investigation of power-amplifier klystrons in connection with their possible modulation characteristics. At the moment we are only concerned with the radio-frequency response, and in this respect attention will now be directed to an important development.

Klystron amplifiers may be constructed with three or even more resonators. These are placed on a common axis as shown in Figure 38 and the beam passes through each in turn.

The resonator nearest to the cathode, R_1, is the initial "buncher" and it is into this resonator that the radio-frequency power is fed in the usual manner. The beam, partially bunched, produces larger voltage swings in the unloaded resonator, R_2. Thus R_2 and R_1 together form a voltage amplifier in which the only power losses are in the resonators. The beam with a new velocity modulation impressed upon it at the gap of R_2 is very thoroughly bunched when it reaches the gap of R_3, the "catcher" resonator, where the radio-frequency power is abstracted in the usual manner. The overall power gain of such a three-resonator klystron is much greater than can be achieved in the corresponding two-resonator amplifier. As has been stated, a gain of 10 db is a fair figure for the latter, while a gain of the order of 25 db is characteristic of the former.

The theory of the three-resonator klystron is complex, for the bunching due to the gap of R_2 is out of phase with the bunching due to R_1, and the radio-frequency current at the gap of R_3 is a complicated function of time. But the practical result is excellent, and when such a device is properly adjusted, the gain is remarkably independent of the input power over a wide range of the latter. The band-width is principally determined by the Q of the second resonator, R_2. By loading this circuit to the requisite extent, the response can be made flat enough for most communication purposes without otherwise affecting the performance.

Three-resonator klystrons can also be used as combined oscillators and buffer amplifiers, but the design of such a valve is quite different from that required for a voltage and power amplifier. In the oscillator R_1 and R_2 are coupled together and that section of the valve is working efficiently when the voltage swing across R_1 is sufficient to produce the most complete bunching at R_2. This means that if the drift space between R_2 and R_3 is long, the beam is over-bunched when it reaches the gap of the latter, and for this reason the drift space between R_2 and R_3 in an oscillator-amplifier is made as short as possible. In a voltage-power amplifier the two drift spaces should be of the same order of magnitude.

3.2.3. *Klystron Frequency Multipliers.* As was pointed out in 2.4.1.6, the current at the "catcher" gap of a klystron is extremely rich in harmonics, and this suggests the use of an amplifier type of valve as a frequency multiplier. The current amplitude of the n^{th} harmonic, expressed as a percentage of the fundamental, may be derived from a table of Bessel Functions. This is shown in Table VI,

TABLE VI

n	1	2	3	4	8	16	24
$\frac{\lvert I_n \rvert}{\lvert I_1 \rvert} \times 100$	100	84	77	69	64	45	40

where the maximum value of each component has been taken in each case.

The other factor which favours the use of the klystron as a frequency multiplier, particularly in the micro-wave region, is the extremely high Q which can be achieved in the catcher resonator. This makes it possible to enhance the response of the latter to a high harmonic with only a negligible response to a neighbouring unwanted harmonic.

For these reasons klystrons are used to multiply by factors of 10 to 20 with a certain success, but there are several important difficulties. The first is that the high harmonic content of the beam current can be considerably reduced by space-charge effects. The axial self-repulsion of the beam electrons tends to flatten each bunch, and so to diminish the sharpness of the change of electron density which constitutes the high harmonic excitation of the beam. The second difficulty is connected with the first. Since debunching must be avoided, it is necessary to use short drift lengths, and these are, as a consequence, much less than is required to optimize the harmonic bunching for normal input. If the power input is increased by a factor equal to the order of the required harmonic, efficient bunching is re-established, but the overall efficiency of the device in terms of radio-frequency power is thereby diminished. This is, in fact, the procedure adopted in practice. The short drift-tube length, characteristic of the output frequency, is employed, and the input power is stepped up to compensate. In addition to these major difficulties there are two minor troubles encountered in using such devices. Klystron multipliers are much more sensitive than amplifiers to changes in the operating parameters. In particular, changes in beam voltage or input power are reflected in large changes of output. This makes multipliers difficult to set up and to operate from normal power supplies.

CHAPTER IV

RECEIVER INPUT CIRCUITS

by P. E. TRIER, M.A.

4.1. **Introduction.** This chapter will be mainly devoted to the theoretical considerations underlying the design of U.H.F. Receiver Input Stages and Circuits. The practical circuit elements and valves have already been treated in earlier chapters; some of the practical design features arising out of the theoretical conclusions will be discussed in the text.

4.1.1. *General Design Considerations.* The function of the receiver is to transfer the aerial voltage at the selected frequency to the first amplifying stage in such a manner that the highest signal/noise ratio is obtained, to amplify the signal to a suitable level in a number of stages, with demodulation at some point, and finally to pass the demodulated signal to the output. At the same time, signals at other frequencies must be rejected with the maximum possible discrimination.

As will be seen below, the stage gain in U.H.F. signal-frequency amplifiers is necessarily very limited. Furthermore, it is difficult to achieve any appreciable selectivity against closely adjacent frequencies in the signal-frequency stages. It is therefore almost universal practice to use superheterodyne reception, in which the signal is transposed from the radio-frequency to a lower intermediate frequency; and the greater part of the necessary gain and selectivity are obtained in the I.F. amplifier chain (Fig. 39).

The design of the I.F. amplifiers and the demodulation circuits of receivers for communication purposes usually follows conventional lines and will not be further discussed here. The discussion will be confined to the signal-frequency amplifier stages, the frequency converter or mixer, and the local oscillator.

The basic design problem is the decision whether a signal-frequency amplifier can be usefully employed before the mixer, or whether the signal is best applied directly to the mixer from the aerial input. If a valve were a true amplifier, giving an output signal identical with the input except for amplitude, no matter how small the input signal, this question would not arise. As it is, however, a number of causes combine to superimpose a random current on the signal current at any point in the receiver. This gives rise to the hissing noise background in the output of a quiescent receiver, and the random currents and voltages are referred to as fluctuation or noise currents and voltages.

The noise voltages are particularly serious in the early stages of the receiver, where the level of the desired signal may be very small. In fact the noise voltage introduced by the first stage of amplification determines the limiting sensitivity of the receiver; if the equivalent noise level introduced by the first stage exceeds the signal level at the input, no amount of subsequent amplification can raise the signal/noise ratio above unity. It is therefore of the highest importance to ensure that the noise level introduced by the first stage of the receiver is kept to a minimum, and that the gain of this stage

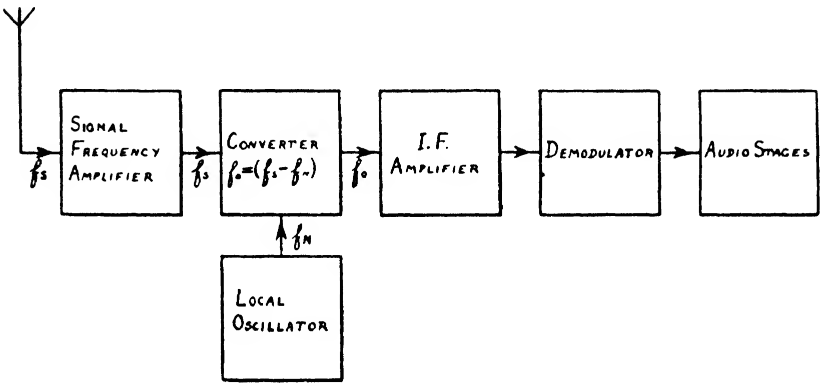


Fig. 39. The block-schematic diagram of a super-heterodyne receiver

is sufficiently high to make the noise contribution from subsequent stages negligible or at least small in comparison. Now it is known that the gain of a valve is always larger when the valve is used as an amplifier than when it is used as a frequency converter stage; on the other hand, the noise contribution is larger when the valve is used as a converter, as will be shown later. It is therefore apparent that at frequencies where real stage gain is possible in an amplifier, a more sensitive receiver will always be obtained if one or several signal-frequency amplifier stages are used before the converter, assuming that the same valve is used for both purposes.

As frequency increases beyond about 200 Mc/s, the efficiency of a pentode or triode converter becomes so low that it is better to use diode or silicon crystal mixers. On the other hand, the conversion efficiency and the noise contribution of a silicon crystal can be held sensibly constant up to frequencies of several thousand megacycles per second. While it is possible to get improved sensitivity by using signal-frequency amplification well beyond the frequency where a valve mixer becomes inefficient, a point will therefore be reached with increasing frequency where the signal-frequency amplifier gain becomes so low, and its noise contribution so high, that no advantage is gained by its use. The point up to which useful gains can be achieved with present-day signal-frequency amplifier valves is in the neighbourhood of 1000 Mc/s. Experimental investigators have claimed advantages from the use of R.F. amplifiers with grounded-grid disc-sealed triodes up to 1500 Mc/s. Above this frequency signal/noise ratio is not improved, and may well be deteriorated, by the use of an R.F. amplifier; and it is best to convert the input signal down to the intermediate frequency before amplifying.

All these remarks apply to R.F. amplifiers using ordinary space-charge modulated valves, i.e. pentodes in the lower frequency ranges up to 200–300 Mc/s and triodes in the higher ranges. Signal-frequency amplifiers using velocity-modulated tubes, travelling-wave tubes, etc., are not yet a practical proposition for receivers, as they are too noisy. It may well be that the limit of useful amplification will be pushed much farther up the spectrum in the future by the application of these principles. In the subsequent treatment, however, consideration will be restricted to pentodes and triodes.

It should be mentioned that it may sometimes be desirable to use signal-frequency amplification in a receiver, even if the sensitivity improvement is insignificant. This occurs when several tuned circuits are necessary at the signal frequency to achieve the image rejection called for; particularly when other considerations preclude the use of a high intermediate frequency.

4.2. Fluctuation Noises in Input Circuits and Valves.

4.2.1. *The Origin of Fluctuation Noises.* The conventional theory of currents in electrical networks is based on the assumption of a continuous current with continuous derivatives. So long as the sensitivity of the measuring device is not driven to an extreme, this assumption works well enough; but with increasing refinement and sensitivity the energy level becomes so small that a point must clearly be reached at which it is no longer possible to disregard the granular or corpuscular nature of electricity, which causes a current to be carried by discrete units of charge, the electrons. It is this granular nature of the current that causes fluctuations in instantaneous intensity round the macroscopic mean value.

These fluctuations determine the limiting values of measurable physical quantities; and it is found quite generally that in any physical measurement, whether electrical or otherwise, the minimum energy supplied to the measuring device must exceed kT, where

$$k = 1{\cdot}38 \times 10^{-23} \text{ joules per degree K}$$

(Boltzmann's Constant)

and T is the absolute temperature in degrees K of the active element of the measuring device.

The quantity kT is a theoretical limit determined thermo-dynamically from the fact that the statistical fluctuations themselves are of this order.

Fluctuation Noise Sources in a Receiver. We shall now examine the fluctuation noise sources in a receiver input circuit. A typical schematic circuit is shown in Figure 40 (*a*), with an equivalent in Figure 40 (*b*).

The following noise sources will be encountered—

1. The aerial will deliver to the circuit, apart from the signal energy, a certain amount of noise energy.

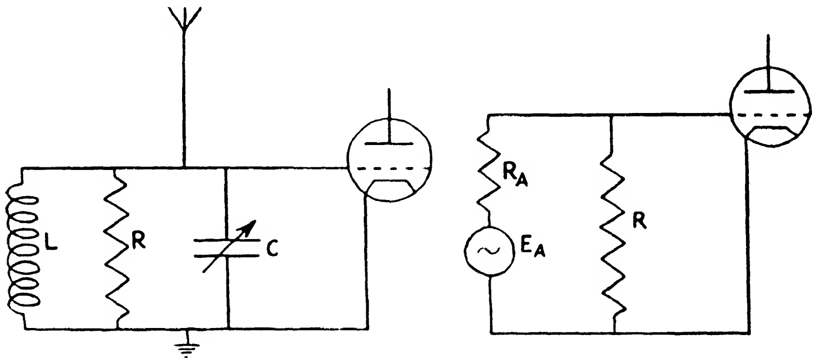


Fig. 40. (*a*) Schematic diagram of a receiver input circuit; (*b*) The equivalent of (*a*) at resonance, showing generator and source impedance

2. The input resistance R of the amplifier will act as a noise source owing to the random thermal agitation of the electrons in the conductor. This noise is described as Thermal or Johnson Noise.

3. The input valve will give rise to a noise because of the random electronic emission from the hot cathode. This is known as Shot Noise. Other more complex noise sources in the valve will also be examined.

Formulae will also be derived below to allow the noise contribution from later stages to be taken into account, but in a well-designed receiver this is small compared with the noise introduced in the first stage.

4.2.2. *Thermal Noise in Conductors.* Thermal noise is best treated before aerial noise is considered. According to the electronic theory of conduction, a conductor is characterized by possessing a large number of free electrons in a state of random motion. The mean kinetic energy of these free electrons is proportional to absolute temperature, and is therefore zero only at 0° K. Each electron in motion constitutes a minute electric current, and the sum total of these currents over a "long" period in the absence of impressed E.M.F.'s is zero. At any given instant, however, these currents do not necessarily add up to zero, and there will be a net current in one direction or the other. This produces a voltage across the ends of the conductor, whose frequency components extend uniformly over the whole spectrum, provided the motion of individual electrons is truly random in the Gaussian sense.

The mean square noise E.M.F., which must be regarded as acting in series with the resistance R of the conductor, has been derived by Nyquist,[1] and is given by

$$\overline{e_n^2} = 4k\text{TBR} \qquad . \quad . \quad . \quad (4.1)$$

where B is the energy band-width of the measuring device or receiver, and T is the absolute temperature of the conductor. For

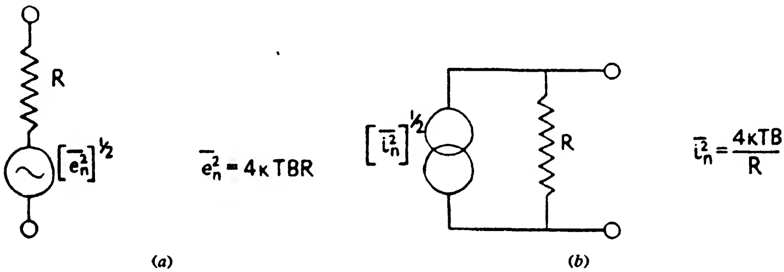


Fig. 41. The equivalent (*a*) constant voltage and (*b*) constant current thermal noise generators

conductors at room temperature, $\text{T} = \text{T}_0 = 290°$ K. Alternatively, the conductor may be regarded as an equivalent current source of mean square value

$$\overline{i_n^2} = \overline{e_n^2}/\text{R}^2 = \frac{4k\text{TB}}{\text{R}} \qquad . \quad . \quad . \quad . \quad (4.2)$$

acting in shunt with the resistance R (Figs. 41A and 41B).

The spectral intensity of the noise power is thus seen to be independent of the frequency, and the total noise power will be directly proportional to the width B of the pass band. Deviations from this law do not occur until we reach exceedingly high frequencies (infra-red), where the Planck energy quantum hf becomes comparable with kT. For all frequencies used in radio, Nyquist's theorem remains valid, although based on classical physical considerations.

If kT is expressed in joules, R in ohms, and B in cycles/second, equations (4.1) and (4.2) give the noise voltages and currents in volts and amperes respectively.

The resistance R in equations (4.1) and (4.2) is the actual or the equivalent resistance of the circuit; if the circuit under consideration is a shunt-resonant tuned circuit, R is therefore the dynamic resistance of the circuit when considered at its resonant frequency; at other frequencies R will be the resistive component of the impedance. In the following treatment, tuned circuits will always be regarded as being at shunt resonance. In circuit problems, the important quantity is often not the open-circuit noise E.M.F., but the noise power available at the terminals for external dissipation. If we imagine a noise-free resistance R_1 connected as a load across the terminals of R, the noise power dissipated in R_1 will be a maximum when $R_1 = R$. This maximum power available for external dissipation is called briefly the available noise power W_n, and we have

$$W_n = \frac{\overline{e_n^2}}{4R} = kTB \quad . \quad . \quad . \quad . \quad (4.3)$$

The available thermal noise power from a conductor is therefore independent of its resistance R.

4.2.3. *Aerial Noise.* Aerial noise must be dismissed briefly here as we are primarily concerned with the receiver itself. For an aerial at thermal equilibrium in an enclosure at temperature T, the mean square noise E.M.F. is again given by Nyquist's equation, which now becomes

$$\overline{e_a^2} = 4kTB(R_0 + R_r) \quad . \quad . \quad . \quad (4.4)$$

where R_0 is the loss resistance, R_r the radiation resistance, and B the energy band-width of the spectrum received.

At frequencies below the U.H.F. region, this thermal equilibrium cannot generally be attained, and the effective temperatures T_0 and T_r of the ohmic and radiation resistances will not be equal, so that equation (4.4) is replaced by

$$\begin{aligned}\overline{e_a^2} &= 4kB(R_0T_0 + R_rT_r)\\ &= 4kBR_aT_a \quad . \quad . \quad . \quad . \quad . \quad (4.5)\end{aligned}$$

where $R_a = R_0 + R_r$, and $T_a = (R_0T_0 + R_rT_r)/(R_0 + R_r)$.

We must therefore regard an aerial in general as the combination of two separate noise sources. The noise due to the loss resistance R_0 is the ordinary thermal or Johnson noise; the noise associated with the radiation resistance is due to the absorption of energy by the aerial from the external radiation field at temperature T_r; and this radiation temperature T_r may differ widely from the ambient temperature T_0, assumed to be 290° K. At frequencies between 20 and 30 Mc/s, measured values of noise level are found to correspond to radiation temperatures T_r ranging from 10 T_0 to about 400 T_0 with a periodic fluctuation corresponding to the sidereal day. This points to a source of very intensive noise radiation in space, the primary maximum of which has recently been localized in the constellation of Sagittarius, the most dense region of the Milky Way. It is doubtful whether the source of this noise is thermal; if it were, it would correspond to temperatures well in excess of 100,000° K, and possibly much higher.

In the U.H.F. region, galactic noise has not the same quantitative effect on aerial radiation temperature; but Hey[(2)] and others have observed serious deterioration in the sensitivity of receivers due to solar noise at 60 Mc/s and above.

4.2.4. *Valve Noise.* In every electronic emission from a hot cathode, the steady current must be accompanied by a fluctuation current caused by the random emission of the electrons; this phenomenon is known as *shot effect*. The emission of the electrons will be truly random if the valve is temperature-limited or saturated, i.e. if there is no space-charge limitation of the emission current, and each emission is an independent process. For a saturated diode, the shot effect can therefore be calculated from first principles; if the emission is limited by space-charge retardation, the theory becomes much more complex.

4.2.4.1. *Shot Effect in a Diode* ([(3)], [(4)]). Consider first a saturated diode, in which each electron is emitted independently of the others. Let $\overline{n_t}$ be the mean number of electrons emitted in an interval t. Then the number n_t of electrons emitted in a particular interval t will in general differ from $\overline{n_t}$. Now the mean square deviation is governed by the following statistical relation

$$\overline{(n_t - \overline{n_t})^2} = \overline{n_t} \quad . \quad . \quad . \quad . \quad (4.6)$$

In other words, the mean square deviation equals the mean value itself; this is a general law for time-counts of random events.

Now let e be the electronic charge, and let I_a be the mean emission current; we then have for the mean square fluctuation current for sample intervals of duration t the following expression

$$\overline{i_a^2} = e^2 \frac{\overline{(n_t - \overline{n_t})^2}}{t^2} = \frac{e^2}{t^2}\overline{n_t} = \frac{e}{t}\left(\frac{e\overline{n_t}}{t}\right) = \frac{e}{t} I_a \quad . \quad . \quad (4.7)$$

The mean square fluctuation current during intervals of duration t is thus proportional to the mean current, and inversely proportional to the duration of the intervals over which measurement takes place. This fluctuation may again be analysed into its spectral components by a Fourier transformation, and the result is

$$\overline{i_a^2} = 2eI_aB \quad . \quad . \quad . \quad . \quad . \quad (4.8)$$

where B is the energy band-width of the measuring instrument. Exactly as in the case of thermal noise in conductors, the spectral power intensity is therefore constant over the whole spectrum, and the noise power is directly proportional to band-width.

Equation (4.8) has been tested experimentally, and found universally valid for saturated emission, so long as electron transit-time effects are negligible. From direct measurement of the fluctuation current, the electronic charge e has been determined with considerable accuracy.

When the emission current is space-charge limited, a modification of equation (4.8) becomes necessary. It is found in this case that the shot noise is much smaller. To understand this, consider the potential distribution in a two-electrode valve (Fig. 42).

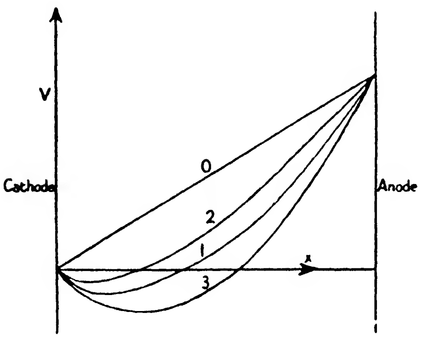


Fig. 42. The mode of variation of the potential relative to the cathode across a diode for varying degrees of space charge

Curve O applies for saturation. If the emission current is space-charge limited, the potential is given by curve 1. The electrons leave the cathode with finite velocities given by a Maxwell distribution, and in front of the cathode, a potential trough is formed where $V < 0$. If the emission fluctuates, the depth and distance of this trough in front of the cathode are varied. Decreasing emission corresponds to curve 2, increased emission to curve 3. In other words, decreased emission will cause a decrease in the space-charge barrier, so that relatively more electrons can get to the anode. This effect is thus seen to exert a smoothing action on the fluctuation current. The shot-effect is space-charge attenuated, and equation (4.8) is replaced by

$$\overline{i_a^2} = F^2 . 2eI_aB \quad . \quad . \quad . \quad . \quad (4.9)$$

where F is the space-charge attenuation factor; $F < 1$. F can be calculated on the basis of the Maxwell distribution of the emission velocities, and the final result can be put in the following equivalent form.

The noise of the space-charge limited diode is equivalent to the thermal noise of a resistance $R_i = \frac{\partial V_a}{\partial I_a}$, where V_a is the anode potential, at a temperature equal to 0·64 times the cathode temperature.

$$\overline{i_a^2} = \frac{0{\cdot}64 \times 4kT_{cath}B}{R_i} \quad . \quad . \quad . \quad . \quad (4.10)$$

Experimental tests of this theoretical formula show less satisfactory agreement than those of formula (4.8) for the saturated diode. Several complicating factors arise in practice; *inter alia* the reflection of electrons from the anode back into the inter-electrode space. The main value of equation (4.10) is its applicability, with one modification, to the triode valve.

4.2.4.2. *Noise Sources in Amplifier Valves.*

Shot Effect. If we replace the diode conductance $1/R_i$ by the mutual conductance g of the triode, the mean square shot noise current of the triode will be given by

$$\overline{i_a^2} = 0{\cdot}64 \times 4kT_{\text{cath}}\,gB \quad . \quad . \quad . \quad (4.11)$$

Partition Noise ([5]). The case of a multi-grid valve is more complex. The fluctuation current is always stronger than that given for a

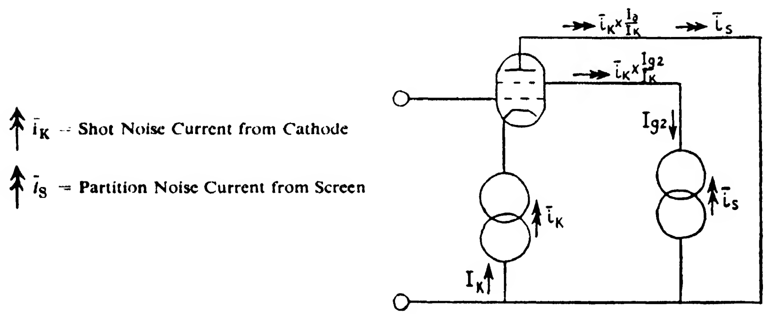


Fig. 43. Noise generators in a tetrode, or in a pentode in which there is a negligible suppressor grid current

triode in equation (4.11). The additional noise current is explained by the fact that the electrons after passing through the negative control grid will take various paths, some of them hitting the screen or suppressor grid, others going straight through to the anode. This random behaviour results in an additional noise current, which is kept down to a minimum in a beam tetrode where the control and screen grids are aligned so that the control grid exerts a shadow effect on the screen grid, and the screen current is kept low. The mean square value of the additional noise current is given by

$$\overline{i_s^2} = 2e\,\frac{I_{g2}I_a}{I_k}\,B \quad . \quad . \quad . \quad . \quad (4.12)$$

where I_a is the anode current, I_{g2} the screen current, I_k the total cathode current. We therefore have for a tetrode or pentode with negligible suppressor current the equivalent diagram of Figure 43.

The inflowing noise current at the cathode is $\bar{i}_k$, caused by shot effect, of which a fraction I_{g2}/I_k goes to the screen, a fraction I_a/I_k to the anode.

At the screen grid we have a further noise current inflow $\overline{i_s}$ due to the partition fluctuations, which goes to the anode. At the anode, both noise currents superimpose. As they are incoherent (independent) noise sources, we must add their mean powers to get the resultant power. This gives for the mean square noise current inflow at the anode

$$\overline{i_a^2} = \overline{i_k^2} \times (I_a/I_k)^2 + \overline{i_s^2}$$
$$= [0{\cdot}64 \times 4kT_{\text{cath}}\, g(I_a/I_k)^2 + 2eI_{g2}I_a/I_k]\, B \qquad . \quad (4.13)$$

In (4.13), the first term gives the shot effect, the second term the partition noise.

The partition noise in a tetrode or pentode becomes particularly harmful at U.H.F. if there is an impedance between cathode and ground, such as a lead inductance, which is difficult to eliminate. This impedance will cause negative feed-back for signal and shot noise currents, i.e. reduce the effective signal input to the valve, but will leave the partition noise current quite unaffected, since it does not flow in the cathode circuit. This is seen from Figure 43, and explains the rapid deterioration of tetrode and pentode performance above 100 Mc/s.

On the other hand, there is the possibility, apparent from Figure 43, of deliberately feeding back the partition noise component from the screen to the grid in such a manner that it should be almost totally smoothed out from the anode current. This was pointed out by Strutt and van der Ziel(6) as early as 1941. The feed-back needed, however, is very critical, and according to Beers(7) it has not been possible to manufacture amplifiers in quantity in which the noise figure was consistently improved by this kind of feed-back. For this reason the detailed analysis will not be given here.

Altogether it should be realized that the real noise behaviour of tetrodes and pentodes is very complex owing to the multiple feed-back paths and the resonances between electrode capacities and lead inductances. The most reliable data on pentode noise at U.H.F. are still obtained from experimental observation rather than from theoretical prediction.

4.2.4.3. *Comparison of Valve Noise Performance.* A mere knowledge of the fluctuation current in the plate circuit of a valve does not enable us to compare the merits of different valves. A valve with a larger noise current may at the same time give a larger gain for the signal input voltage, so that the signal/noise ratio at the output may still be superior to that of another valve with less inherent noise. We must therefore refer the plate circuit noise back to a fictitious equivalent noise source in the grid input circuit of the valve, to allow a comparison of the merits of different valves. The most useful concept by which this can be carried out is to specify the resistance, which when applied between the control grid and cathode at room temperature ($T_0 = 290°$ K) would give the same fluctuation current in the plate circuit (after noiseless amplification in the valve) as the shot and partition noises give in actual fact. This fictitious resistance is called the *equivalent noise resistance* R_{eq} of the valve, and affords a direct figure of merit.

If the total mean square noise current in the plate circuit is $\overline{i_a^2}$, then the equivalent noise resistance is, by the above definition,

$$R_{eq} = \frac{\overline{i_a^2}}{g^2 \cdot 4kT_0B} \qquad (4.14)$$

For a given valve, R_{eq} can be calculated by combining equations (4.13) and (4.14). If we take the normal oxide-coated cathode ($T_{cath} = 1150°$ K), currents in mA, mutual conductances in mA/V, we obtain for the equivalent noise resistance in kilo-ohms

$$R_{eq} = \frac{2 \cdot 5}{g}\left(\frac{I_a}{I_k}\right)^2 + \frac{20}{g^2}\frac{I_a I_{g2}}{I_k} \qquad (4.15)$$

or for a triode simply

$$R_{eq} = \frac{2 \cdot 5}{g} \qquad (4.15a)$$

The equivalent noise resistance can thus be derived from the static valve characteristics, and agreement with direct noise measurement is close.

It is seen that the demand for a low equivalent noise resistance calls for a low screen current I_{g2}, and for a high mutual conductance, g.

Typical equivalent noise resistance values are—

Input triodes, 300–2000 ohms.
High slope pentodes, 700–5000 ohms.
Remote cut-off pentodes, 10,000–50,000 ohms.
Multigrid frequency changers, up to 100,000 ohms.

It must be stressed that the equivalent noise resistance R_{eq} is purely fictitious, and is represented in the circuit simply by a noise E.M.F. in the grid lead of mean square value

$$\overline{e_v^2} = 4kT_0R_{eq}B \qquad (4.16)$$

4.2.4.4. *Induced Noise: Transit-time Loading*. It has already been pointed out in Chapter II that several effects combine to increase the input admittance of a valve at high frequencies. The first effect to become noticeable with increasing frequency is due to the cathode lead inductance L_k, which in conjunction with the grid-cathode capacity C_{gk} causes a resistive admittance

$$Y_L = g\omega^2 L_k C_{gk} \quad . \quad . \quad . \quad . \quad (4.17)$$

where ω is the angular frequency and g the mutual conductance.

This admittance may be imagined replaced by an external resistance across the grid-cathode circuit, and must be taken into account when calculating the thermal noise from the external circuit. If it is regarded purely as a feed-back effect, its noise temperature should be zero, as will be shown below in the discussion of grounded-grid circuits; but, in practice, more consistent results are obtained by putting its noise temperature $\sim T_0$.

The second effect to increase the input admittance as frequency increases is the finite time of transit of the electrons between cathode and grid. If this transit-time is no longer negligible compared with the period of the impressed signal, the grid charge will go through its maximum value later than the grid voltage. This lag in the grid charge will result in a corresponding lag of the (capacitative) grid current; grid voltage and current will no longer be in quadrature and there will be a real component in the input admittance due to this cause. For small transit angles, the resistive input conductance is given by

$$Y_t = Ag\omega^2 t^2 \quad . \quad . \quad . \quad . \quad (4.18)$$

where A is a constant and t is the cathode-grid transit-time.[8,9] This input admittance may again be replaced by an equivalent external resistance across the input of the valve. Transit-time loading again gives rise to a fluctuation noise current, since the fluctuations in the emission current induce a noise current in the grid, which in turn reacts on the electron stream across the tube. North and Ferris[10] show that the induced noise current caused by transit-time loading is equal to the thermal noise produced by the equivalent resistance across the input, provided this resistance is assumed to be at a temperature $\sim 5T_0$. This noise current is not strictly incoherent with the shot noise, but since the coherence is a quadrature relationship, the induced noise may be treated as incoherent with the shot noise for small transit angles, and hence its mean square value added to that of the other noise currents in the valve.

In practice it is impossible to separate the effects given by equations (4.17) and (4.18), since the admittance in both cases varies directly as g and as ω^2. We therefore have a shunt resistance across the valve input, part of which is at noise temperature T_0 or less, part at $5T_0$. In addition, there is a further noise source due to total emission damping at high frequencies, the effect of which has not yet been clearly analysed. Finally, there is the resistance of the cathode coating itself, the effect of which is only now beginning to be understood. It is therefore usual to combine all these effects into a single shunt resistance across the input at a temperature βT,[11] where β has some value between 1 and 5. Values for β less than 1 appear theoretically possible, but are not found in practice. In calculations, a value $\beta = 1$ is usually adequate; but for disc-seal valves in U.H.F. circuits, the lead inductance effect is usually negligible

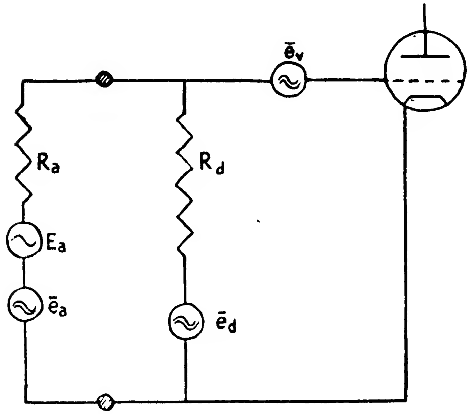


Fig. 44. Detailed diagram of a receiver input circuit, showing the various noise generators

compared with the transit-time effect, and we may put $\beta = 5$. The factor β is called the *Noise Temperature Ratio* of the valve input resistance. The uncertainty in β does not critically affect signal/noise ratio calculations, so long as matching for optimum noise rather than for maximum power transfer is used in the input circuit. This will be shown below.

4.2.5. *Calculation of Signal/Noise Ratios.*

Noise Factor. After the discussion of the various sources of noise in an input circuit, we are now in a position to draw a more complete equivalent diagram of the circuit in Figure 40A with the noise sources indicated; and from this we can derive the signal/noise ratio at the output of the stage for a given input signal/noise ratio (Fig. 44).

E_a is the R.M.S. open-circuit aerial signal E.M.F., R_a the source impedance, and $\bar{e}_a$ the R.M.S. aerial noise E.M.F. The shunt impedance in the input circuit is R, as in Figure 40, but since the circuit losses are usually quite negligible compared with the valve input loss, we may put $R = R_d$, where R_d is the high-frequency valve input impedance. The R.M.S. noise voltage due to this input impedance is $\bar{e}_d$, and the equivalent R.M.S. shot noise voltage at the grid is $\bar{e}_v$.

The noise due to the plate circuit load impedance will be neglected, since its transferred value in the grid circuit is small, and its magnitude is in any case negligible compared with the noise contribution from the following stage, which will be considered later.

The R.M.S. signal voltage E_g at the grid will be

$$E_g = E_a \frac{R_d}{R_d + R_a} \quad . \quad . \quad . \quad (4.19)$$

The effects of the separate noise voltages at the grid are

Aerial noise $\overline{e_a} \dfrac{R_d}{R_a + R_d}$

where $\overline{e_a^2} = 4kT_aR_aB = 4k\alpha T_0R_aB$, α being the aerial noise temperature ratio.

Input circuit noise $\overline{e_d} \dfrac{R_a}{R_a + R_d}$

where $\overline{e_d^2} = 4k\beta T_0R_dB$, β being the noise temperature ratio of the valve input impedance.

Valve noise $\bar{e}_v$

where $\overline{e_v^2} = 4kT_0R_{eq}B$.

Since these are all incoherent noise sources, we must add their mean square voltages to get the resultant mean square noise voltage $\overline{e_g^2}$ at the grid.

$$\overline{e_g^2} = \overline{e_a^2} \frac{R_d^2}{(R_a + R_d)^2} + \overline{e_d^2} \frac{R_a^2}{(R_a + R_d)^2} + \overline{e_v^2} \quad . \quad . \quad . \quad (4.20)$$

$$= 4kT_0B \left[\alpha R_a \frac{R_d^2}{(R_a + R_d)^2} + \beta R_d \frac{R_a^2}{(R_a + R_d)^2} + R_{eq} \right] \quad . \quad (4.21)$$

From equations (4.19) and (4.21) we can calculate the output signal/noise power ratio $E_g^2/\overline{e_g^2}$, if the input signal/noise ratio $E_a^2/\overline{e_a^2}$ is known. Note that $E_g^2/\overline{e_g^2}$ represents the signal/noise power ratio at the output of the stage, although both signal and noise have been referred back to the grid as explained in Section 4.2.4.3.

The more usual question, however, is to find the input signal/noise ratio necessary to give unity signal/noise ratio at the output, as this will obviously provide a figure of merit for the sensitivity of the stage, viewed as a four-terminal network.

From equations (4.19) and (4.21), we have for the input signal/noise ratio

$$\frac{E_a^2}{\overline{e_a^2}} = \frac{E_g^2}{\overline{e_g^2}} \left[1 + \frac{\beta R_a}{\alpha R_d} + \frac{(R_a + R_d)^2}{\alpha R_a R_d^2} R_{eq} \right] \quad . \quad . \quad (4.22)$$

In this expression, the aerial noise temperature ratio α enters. In order to get a figure of merit for the input stage apart from the thermal nature of the external source, we specify an equivalent source of internal impedance R_a, such as a signal generator at room temperature $T_0 = 290°$ K, making $\alpha = 1$.

The input signal/noise ratio now becomes

$$\frac{\overline{E_a^2}}{\overline{e_a^2}} = \frac{\overline{E_g^2}}{\overline{e_g^2}}\left[1 + \beta\frac{R_a}{R_d} + \left(\sqrt{\frac{R_a}{R_d}} + \sqrt{\frac{R_d}{R_a}}\right)^2 \frac{R_{eq}}{R_d}\right]. \quad (4.23)$$

or

$$\frac{\overline{E_a^2}}{\overline{e_a^2}} = \eta\frac{\overline{E_g^2}}{\overline{e_g^2}} \quad (4.24)$$

where

$$\eta = 1 + \beta\frac{R_a}{R_d} + \left(\sqrt{\frac{R_a}{R_d}} + \sqrt{\frac{R_d}{R_a}}\right)^2 \frac{R_{eq}}{R_d} \quad (4.25)$$

The factor η is defined as the *noise factor* of the stage for a source impedance R_a; it may be regarded as the ratio of the signal/noise power ratios at the input and at the output of the four-terminal network, or more simply as the input signal/noise power ratio necessary to give unity signal/noise ratio at the output, always provided that the noise at the input is simply the thermal noise corresponding to source impedance R_a.

For unity signal/noise ratio at the output, we therefore have

$$E_a^2 = \eta\,\overline{e_a^2} = \eta 4kT_0R_aB$$

Now the available signal power from the source is

$$W_a = E_a^2/4R_a = \eta kT_0B = \eta W_n \quad (4.26)$$

where W_n is the available noise power from the source, from equation (4.3).

We may therefore define the noise factor as the factor by which the available signal power from the source must exceed the available noise power from the source to achieve unity signal/noise ratio at the output. Since this definition of the noise factor is a power ratio, the noise factor is often expressed in decibels for practical purposes. For theoretical developments, however, it is better left as a direct numerical ratio, and will be so expressed throughout the following treatment.

4.2.5.1. *Noise Factor for Several Stages in Cascade.* Consider a source giving available signal power S_0, and available noise power N_0 ($= kT_0B$) (Fig. 45). Connect to a stage of power gain G_1, noise factor η_1, and a further stage of power gain G_2, noise factor η_2, in cascade. From the definition of noise factor, we may regard the first stage as contributing a further noise power $(\eta_1 - 1)N_0$ before noiseless amplification; the second stage contributes noise power $(\eta_2 - 1)N_0$ before noiseless amplification.

Let the signal and noise powers available at the output of the first stage be S_1 and N_1 respectively; the signal and noise powers available at the output of the second stage S_2 and N_2 respectively. Then the composite noise factor will be the input signal/noise power ratio required to give unity output signal/noise ratio S_2/N_2.

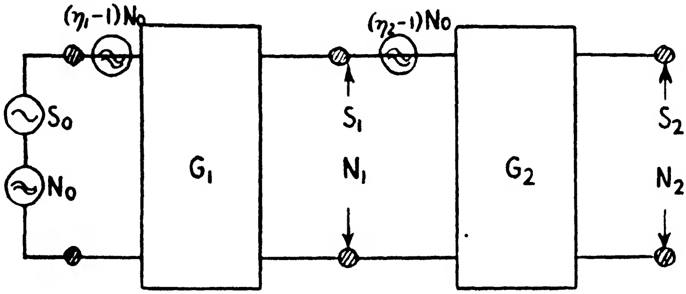


Fig. 45. Block diagram of two amplifier stages in cascade, showing the relations of the signal and noise generators

Since amplification itself is assumed noiseless, we have

$$\frac{S_2}{N_2} = \frac{S_1}{N_1 + (\eta_2 - 1)N_0}$$
$$= \frac{G_1 S_0}{G_1 \eta_1 N_0 + (\eta_2 - 1)N_0}$$
$$= \frac{S_0}{N_0} \cdot \frac{1}{\eta_1 + \frac{\eta_2 - 1}{G_1}} \quad . \quad . \quad . \quad . \quad (4.27)$$

Therefore, for unity output signal/noise ratio S_2/N_2, we have for the composite noise factor of the two stages in cascade

$$\eta_{1+2} = \frac{S_0}{N_0} = \eta_1 + \frac{\eta_2 - 1}{G_1} \quad . \quad . \quad . \quad (4.28)$$

Equation (4.28) is easily extended by induction to the case of more than two stages, e.g. for three stages

$$\eta_{1+2+3} = \eta_{1+2} + \frac{\eta_3 - 1}{G_1 G_2}$$
$$= \eta_1 + \frac{\eta_2 - 1}{G_1} + \frac{\eta_3 - 1}{G_1 G_2} \quad . \quad . \quad (4.29)$$

Equations (4.28) and (4.29) are very important, as they afford a criterion whether any useful purpose can be served by R.F. amplifier stages. Normally, if the gain G_1 of the first stage is considerable, the noise contributions of the subsequent stages may be disregarded, and the composite noise factor is virtually equal to that of the first stage alone. As the gain decreases with increasing frequency, noise from later stages becomes considerable and must be taken into account. A point will be reached, and can be determined from (4.28), if the gain is known, where the addition of an R.F. stage in front of the mixer no longer gives any improvement in the noise factor.

4.2.5.2. *Measurement of Noise Factor*. Noise factor measurement is most easily carried out by using a noise generator giving a known noise power output per unit band-width. The saturated diode is such a source. A directly heated diode with a pure tungsten filament is used for this purpose. From equation (4.8), there will be a noise component in the anode current of mean square value $\overline{i_a^2} = 2eI_aB$, where I_a is the D.C. anode current.

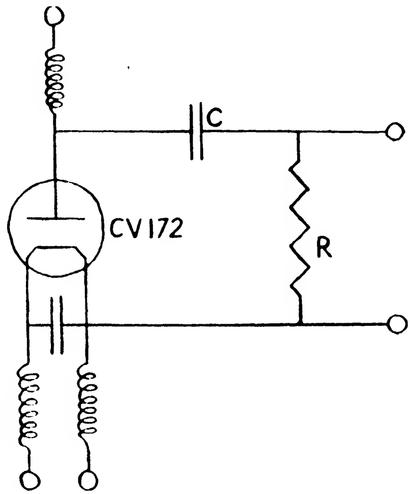


Fig. 46. Diode noise generator

If a resistance R is connected across the diode through a small condenser C, as in Figure 46, the diode noise current will set up a noise voltage across the resistor of mean square value

$$\overline{e_s^2} = R^2 . \overline{i_a^2} = R^2 . 2eI_aB \qquad . \quad . \ (4.30)$$

provided R is small compared with the diode impedance. In addition to this noise voltage, there will also be the thermal noise voltage across R, of mean square value

$$\overline{e_r^2} = 4kT_0RB$$

We may now connect R across the input of the receiver under test, and regard $\overline{e_s^2}$ as the "signal," and $\overline{e_r^2}$ as the "noise" at the receiver input. The noise power at the receiver output is measured. If P is the ratio of the noise power output of the receiver when the noise diode is switched on, to the noise power output when the diode filament is switched off, then the "signal"/noise ratio at the output is $P - 1$.

Hence the noise factor of the receiver for a source impedance R is

$$\eta = \frac{\overline{e_s^2}}{\overline{e_r^2}} \Big/ (P - 1) = \frac{R^2 \, . \, 2eI_aB}{4kT_0RB(P-1)} = \frac{ReI_a}{2kT_0(P-1)} \quad . \quad (4.31)$$

Inserting the numerical values for e and kT_0, this becomes

$$\eta = \frac{20RI_a}{P-1} \quad . \quad . \quad . \quad . \quad . \quad (4.32)$$

where R is in ohms, and the diode current I_a in amps. R should, of course, be chosen in accordance with the source impedance specified for the receiver.

For measurements in the U.H.F. field, precautions must be taken to minimize the effect of the lead inductances between the electrodes of the diode and the resistance R, which may falsify the results. Special noise diodes are in use, in which the electrode structure is built integrally into a coaxial transmission line, terminated at the end by the correctly matched resistance R.

The noise output measurement should strictly speaking be made at the output of the I.F. amplifier, before demodulation. In this case, a thermocouple or bolometer must be used for the noise power measurement, with a uniform frequency response over the whole pass-band of the I.F. amplifier.

The noise power measurement may be made directly at the audio output of the receiver if the action of the second detector is linear. Since this is usually not a permissible assumption when the level is low, the second detector action may be artificially linearized by the injection of a sufficiently large unmodulated carrier signal at the I.F. mid-frequency before the detector.

4.3. Design of Signal-frequency Amplifier Stages.

Noise Factor, Gain and Band-width Considerations. In Section 4.2.5, the signal/noise ratio of a simple input amplifier stage was calculated in terms of the source impedance, the circuit impedance and the equivalent noise resistance of the valve, and a definition of noise factor was derived. It is now necessary to examine the conditions for matching the source to the circuit which will lead to an optimum value for the noise factor, and at the same time to take selectivity considerations into account. This will be done first for a grid-input (grounded-cathode) circuit stage, then for a cathode-input (grounded-grid) stage.

4.3.1. *The Grid-input Amplifier*

4.3.1.1. *Determination of Optimum Noise Factor.*

Consider the input stage in Figure 47A. The tuned circuit is shown schematically as a conventional L-C circuit, but will of course in practice usually be a butterfly circuit, a resonant line, or even a resonant cavity. The aerial input is tapped down on the circuit with a step-up ratio from the aerial of 1 : m, i.e. the aerial voltage will be stepped up by a factor m across the tuned circuit, and the aerial impedance, as seen by the tuned circuit, will be multiplied by m^2. The losses in the tuned circuit are neglected as before, and the circuit shunt impedance is therefore represented by the valve input impedance R_d. The problem is to determine the step-up ratio m which will give the minimum noise factor, and also to determine the value of this minimum noise factor.

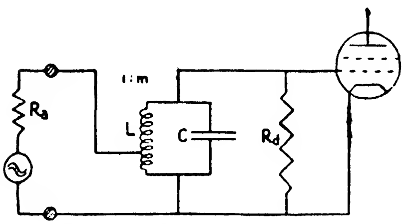


Fig. 47A. Receiver input stage employing grounded-cathode valve

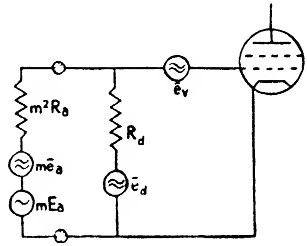


Fig. 47B. Equivalent circuit of stage shown in Fig. 47A

In Figure 47B, the R.M.S. noise voltages in the circuit are inserted, and their mean square values will be given by

$$\overline{e_a^2} = 4kT_0R_aB \quad . \quad . \quad . \quad (4.33)$$

(The source noise temperature will now be taken to be T_0, to derive a standardized noise figure.)

$$\overline{e_d^2} = 4k\beta T_0R_dB \quad . \quad . \quad . \quad (4.34)$$

$$\overline{e_v^2} = 4kT_0R_{eq}B \quad . \quad . \quad . \quad (4.35)$$

Now let

$$R_d = p^2R_a \quad . \quad . \quad . \quad (4.36)$$

so that p is the step-up ratio corresponding to the maximum power transfer from the aerial to the circuit.

We have, for the R.M.S. signal voltage E_g at the grid

$$E_g = mE_a\frac{R_d}{R_d + m^2R_a} = mE_a\frac{p^2}{p^2 + m^2} \quad . \quad . \quad (4.37)$$

The effects of the several noise voltages at the grid will be

$$\overline{e_1} = m\overline{e_a}\frac{R_d}{R_d + m^2R_a} = m\overline{e_a}\frac{p^2}{p^2 + m^2} \quad . \quad . \quad (4.38)$$

$$\overline{e_2} = \overline{e_d}\frac{m^2R_a}{R_d + m^2R_a} = \overline{e_d}\frac{m^2}{p^2 + m^2} \quad . \quad . \quad (4.39)$$

$$\overline{e_3} = \overline{e_v} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.40)$$

The mean square noise voltage $\overline{e_g^2}$ at the grid will therefore be

$$\overline{e_g^2} = \overline{e_1^2} + \overline{e_2^2} + \overline{e_3^2} \quad . \quad . \quad . \quad (4.41)$$

From equations (4.33) to (4.41) we can now derive the noise factor,

$$\eta = \frac{E_a^2}{\overline{e_a^2}} \Big/ \frac{E_g^2}{\overline{e_g^2}} = 1 + \beta \frac{m^2}{p^2} + \left(\frac{p}{m} + \frac{m}{p}\right)^2 \frac{R_{eq}}{R_d} \quad . \quad (4.42)$$

The value m_0 of the step-up ratio m which makes the noise factor a minimum is found by equating the term in m^2 to the term in $1/m^2$ in equation (4.42). This gives

$$\left(\beta + \frac{R_{eq}}{R_d}\right)\frac{m^2}{p^2} = \frac{R_{eq}}{R_d}\frac{p^2}{m^2}$$

or

$$\frac{m_0^4}{p^4} = \frac{R_{eq}}{R_{eq} + \beta R_d} \quad . \quad . \quad (4.43)$$

Equation (4.43) gives the step-up ratio $m = m_0$ for optimum noise factor, as a fraction of the "maximum power transfer" step-up ratio p. It is seen that "matched" transfer from aerial to circuit will *not* in general give optimum noise factor, except at the very highest frequencies, where the input impedance R_d of the valve may be small compared with the equivalent noise resistance R_{eq}. The optimum noise factor corresponds to a step-up ratio m_0 *smaller* than the matched transfer ratio p, i.e. the aerial should be more tightly coupled to the circuit than for the matched condition. Moreover, since the noise temperature ratio β of the valve input impedance enters critically into equation (4.43), and this may not be accurately known, the matching for optimum noise factor must in practice be carried out by trial. The optimum is however not very critical,[11] as equation (4.43) shows; even if the correct value for m_0^4/p^4 is considerably smaller than 1, m_0/p may still be nearly unity.

Finally, inserting the optimum value for m/p from equation (4.43) in equation (4.42), the optimum noise factor itself becomes

$$\eta_{OPT} = 1 + 2\left[\frac{R_{eq}}{R_d} + \sqrt{\left(\frac{R_{eq}}{R_d}\right)^2 + \beta\frac{R_{eq}}{R_d}}\right] \quad . \quad . \quad (4.44)$$

The ratio R_{eq}/R_d must therefore be made as small as possible. We must always use valves with the lowest possible equivalent noise resistance R_{eq}, and the highest possible input impedance R_d. Pentodes will therefore give an inherently higher noise factor than input triodes, since their noise resistance R_{eq} has been shown to be higher. On the other hand, if a grid input circuit is used with a triode, neutralization will usually be required to prevent self-oscillation.

Since the input impedance R_d of a valve varies inversely as the square of the frequency, R_d decreases rapidly with increasing frequency, and the performance of the valve falls off. We have seen that the two main causes of this behaviour are the lead inductances and transit-time effects. Lead inductances are minimized by using disc-seal techniques; transit-time is minimized by the closest possible cathode-grid spacing, and by increasing the emission current density per unit area to a maximum.

The noise factor for the "matched power transfer" condition follows from equation (4.42) by putting $m/p = 1$.

$$\eta_{\text{matched}} = 1 + \beta + 4\frac{R_{eq}}{R_d} \quad . \quad . \quad . \quad (4.45)$$

Since $\beta \geq 1$, the noise factor for this condition must always be greater than 2 and it is important to improve it, where necessary, by the correct mismatch given in equation (4.43).

Returning to the case of matching for optimum noise factor, we may write, for any valve,

$$\frac{R_{eq}}{R_d} = \frac{f^2}{f_n^2} \quad . \quad . \quad . \quad . \quad . \quad (4.46)$$

where f is the frequency, and f_n is the frequency at which, for the particular valve type, the valve input impedance R_d is equal to its noise resistance R_{eq}. This relation assumes that the equivalent noise resistance is sensibly constant over the whole useful frequency range of the valve.

For a noise temperature ratio $\beta = 1$, equation (4.44) now becomes

$$\eta_{\text{OPT}} = 1 + 2\left[\frac{f^2}{f_n^2} + \left(\frac{f^2}{f_n^2} + \frac{f^4}{f_n^4}\right)^{\frac{1}{2}}\right] \quad . \quad . \quad . \quad (4.47)$$

giving the optimum noise factor in terms of frequency of operation, and the frequency f_n defined as above. This relation is plotted in Figure 48, and the optimum noise factor is expressed in decibels as a function of frequency. The corresponding curve for a noise temperature ratio $\beta = 5$ is about 1 db worse throughout, up to $f = 2f_n$. Although the noise temperature ratio β critically affects the optimum matching, it therefore does *not* critically affect the optimum noise factor itself.

From Figure 48, the performance of any valve may be assessed if f_n is known. The table shown on p. 110 gives R_{eq} and f_n for some typical valves.

Figure 48 gives optimum noise factors up to $f = 2f_n$. This frequency $f = 2f_n$ may be regarded as the useful upper frequency limit of a signal-frequency amplifier. At higher frequencies, the noise factor begins to compare unfavourably with that of a well-designed crystal mixer, and for pentodes the simple relationships outlined break down owing to internal resonances and multiple feed-back paths.

At the lower frequency end, below $f = 0{\cdot}4f_n$, the curve of Figure 48 tends to give optimistic results when compared with practical

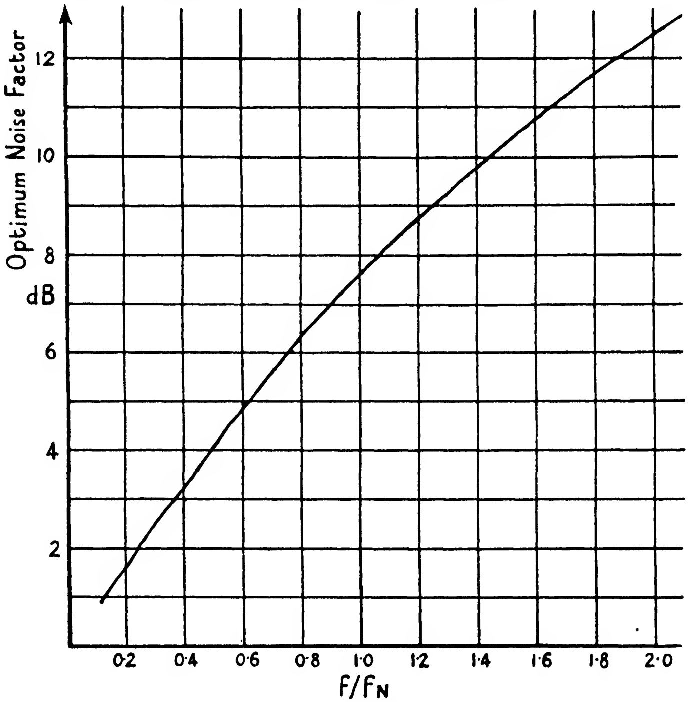


Fig. 48. Universal noise factor characteristic, giving the optimum noise factor theoretically obtainable for any valve in terms of f/f_n. Valve loss only is considered. The curve is for $\beta = 1$. If $\beta = 5$, the noise factor is about 1 db worse throughout

	Type	R_{eq}, Ohms	f_n, Mc/s
Pentodes	EF 50	1400	90
	CV 138	1000	125
	EF 42	750	125
	954	6000	160
	6AK5	2500	160
Triodes	CV 139/6J4	220	300
	CV 53	450	500
	CV 354	400	750

achievements, since the losses in the tuned circuit have been neglected, and these become appreciable at the lowest frequencies in comparison with the decreased valve losses.

4.3.1.2. *The Gain of the Grid-input Stage.* We have already seen that it may be necessary to mismatch the valve to the source. For this reason it is inadequate to measure stage gain simply as the ratio of power output to actual power input into the valve. The power gain of a stage, considered as a four-terminal network, is therefore best defined by the relation

$$G = P_0/P_i \quad . \quad . \quad . \quad . \quad (4.48)$$

where P_0 is the *available* power at the output terminals, and P_i is the *available* power at the input terminals.

Referring to Figure 47B, we have for the available input power

$$P_i = \frac{(mE_a)^2}{4m^2R_a} = \frac{E_a^2}{4R_a} \quad . \quad . \quad . \quad . \quad (4.49)$$

The actual power may of course be different, since it depends on the step-up ratio *m*.

To obtain the available output power, turn to the equivalent plate circuits for triodes and pentodes in Figures 49A and 49B.

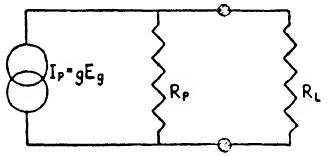


Fig. 49A. Equivalent plate circuit for a grounded-cathode triode

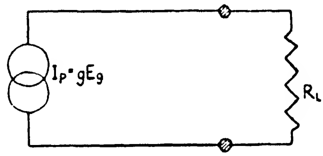


Fig. 49B. Equivalent plate circuit for a pentode

For a triode, the current source $I_p = gE_g$ is in shunt with the internal plate resistance R_p and the load resistance R_L. The power available for dissipation in R_L is therefore

$$P_0 = \frac{(gE_g)^2R_p}{4} \quad . \quad . \quad . \quad . \quad (4.50)$$

For a pentode, the internal plate impedance is very large, and we may neglect its shunting effect on R_L in the equivalent circuit. We therefore have for the pentode

$$P_0 = (gE_g)^2R_L \quad . \quad . \quad . \quad . \quad (4.51)$$

For the gain evaluation it is convenient to introduce the total impedance R_c in the grid-cathode circuit, consisting of the transformed impedance m^2R_a in shunt with the valve input impedance R_d.

$$R_c = \frac{m^2R_aR_d}{m^2R_a + R_d} = R_a\frac{m^2p^2}{m^2 + p^2} \quad . \quad . \quad (4.52)$$

The voltage E_g at the grid is given by

$$\frac{E_g}{mE_a} = \frac{R_d}{m^2R_a + R_d} = \frac{R_c}{m^2R_a} \quad . \quad . \quad (4.53)$$

and the power gain for the triode stage will be

$$G = P_0/P_i$$
$$= g^2R_pR_a\frac{E_g^2}{E_a^2} = g^2\frac{R_pR_c^2}{m^2R_a} \quad . \quad . \quad (4.54)$$

This may be transformed by putting

$$\frac{m^2R_a}{R_d} = \frac{m^2}{p^2} = \lambda^2 \quad . \quad . \quad . \quad . \quad . \quad (4.55)$$

and

$$\frac{R_c}{m^2R_a} = \frac{p^2}{m^2 + p^2} = \frac{1}{1 + \lambda^2} \quad . \quad . \quad (4.56)$$

Then

$$G = g^2R_pR_d\frac{\lambda^2}{(1 + \lambda^2)^2}$$
$$= \frac{g^2R_pR_d}{\lambda^2 + 2 + \frac{1}{\lambda^2}} \quad . \quad . \quad . \quad . \quad . \quad (4.57)$$

The maximum possible gain will clearly be achieved for

$$\lambda = \frac{m}{p} = 1,$$

and will be

$$G_{max} = \frac{g^2R_pR_d}{4} = \frac{g\mu R_d}{4} = \frac{\mu^2R_d}{4R_p} \quad . \quad . \quad (4.58)$$

Equation (4.58) brings out the design features necessary for a triode with a high stage gain; viz. a high mutual conductance g, a high plate impedance R_p, and a high input impedance R_d. Equation (4.58) also shows, since the valve input impedance R_d is inversely proportional to the square of the frequency, that the maximum stage gain is inversely proportional to the square of the frequency, and therefore rapidly falls off with increasing frequency.

The gain obtainable for optimum noise factor is given by substituting in equation (4.57) the value for λ^2 previously determined in equation (4.43)

$$\lambda^2 = \left(\frac{R_{eq}}{R_{eq} + \beta R_d}\right)^{\frac{1}{2}}$$

For a pentode, the relation for the gain will be

$$G = \frac{4g^2R_dR_L}{\lambda^2 + 2 + \frac{1}{\lambda^2}} \quad . \quad . \quad . \quad (4.59)$$

and the maximum gain will be, for $\lambda = 1$

$$G_{max} = g^2R_dR_L \quad . \quad . \quad . \quad (4.60)$$

The gain of the pentode stage can therefore be increased by the use of a higher load resistance R_L, but in the U.H.F. field, a limit is set to this by the input impedance of the following stage.

4.3.1.3. *Band-width Considerations and their Effect on Noise Factor.* The band-width necessary for a signal-frequency amplifier in a receiver is determined by the following factors—

1. The spectrum taken up by the transmitted intelligence.
2. The frequency tolerance of the transmitter carrier.
3. The stability and ease of tuning of the receiver R.F. circuits.
4. The ganging tolerance if several R.F. circuits are to be tuned by a common shaft.

If there is a possibility of interference from strong signals on adjacent frequencies, or if rejection of the I.F. image frequency is important, it is necessary to restrict the band-width of the R.F. amplifier to the minimum consistent with the above factors.

The overall selectivity of the R.F. amplifier can be gauged by reference to the half-power band-width of the individual tuned circuit, and this will therefore be examined here.

Referring to Figure 47A and Figure 47B, the dynamic impedance determining the band-width of the circuit is the total grid-cathode impedance R_c, consisting of R_d in shunt with m^2R_a, neglecting losses in the tuned circuit itself. For the single-tuned circuit, the band-width between half-power points is given by

$$B = \frac{f_0}{Q} = \frac{f_0}{2\pi f_0 CR_c} = \frac{1}{2\pi CR_c} = \frac{1}{2\pi C}\left(\frac{1}{m^2R_a} + \frac{1}{R_d}\right) \quad (4.61)$$

where C is the total equivalent shunt capacitance in the grid-cathode circuit. If the resistances are given in ohms, and C in microfarads, the band-width B will be in Mc/s.

It should be pointed out that the half-power band-width does not correspond exactly to the energy band-width used hitherto, but the difference is sufficiently small to be neglected here. The strict definition of the energy band-width is

$$B = \frac{1}{G_0}\int_0^\infty G\,df$$

where G is the power gain as a function of frequency, and G_0 is the maximum power gain at the mid-band frequency f_0.

Low-frequency Case. Consider first the low-frequency case, where the input conductance $1/R_d$ of the valve is small and may be neglected. The band-width now becomes

$$B = \frac{1}{2\pi C} \cdot \frac{1}{m^2 R_a} \quad . \quad . \quad . \quad . \quad (4.62)$$

Under these conditions, the noise factor may be deduced from (4.42) by substituting $R_d = p^2 R_a$, and then making $1/p^2$ tend to zero. The result is

$$\eta = 1 + \frac{R_{eq}}{m^2 R_a} \quad . \quad . \quad . \quad . \quad (4.63)$$

Combining (4.62) and (4.63), we have for the noise factor at frequencies where $1/R_d$ is negligibly small,

$$\eta = 1 + 2\pi C B R_{eq} \quad . \quad . \quad . \quad (4.64)$$

In this case, the noise factor depends only on the required band-width; if a small band-width is adequate, the noise factor can be reduced very nearly to unity; on the other hand a larger band-width will necessarily deteriorate the noise factor. The band-width itself is here assumed to be entirely dependent on the transformer step-up ratio *m*, since circuit and valve losses are neglected. A higher step-up ratio will reduce both noise factor and band-width.

General Case—Band-width for Optimum Noise Factor. Returning to the general case at higher frequencies, where the valve input conductance $1/R_d$ may no longer be neglected, we shall determine the band-width for the transformer step-up ratio giving optimum noise factor, and its relation to the noise factor. We have

$$2\pi CB = \frac{1}{R_d}\left(1 + \frac{p^2}{m^2}\right) \qquad \text{from (4.61)}$$

$$= \frac{1}{R_d}\left(1 + \sqrt{1 + \beta \frac{R_d}{R_{eq}}}\right) \qquad \text{from (4.43)} \quad . \quad (4.65)$$

giving

$$B_0 = \frac{1}{2\pi C R_d}\left(1 + \sqrt{1 + \beta \frac{R_d}{R_{eq}}}\right) \quad . \quad . \quad (4.66)$$

(Band-width for optimum noise factor step-up.)

From (4.65) we have further

$$4\pi C B_0 R_{eq} = 2\frac{R_{eq}}{R_d}\left(1 + \sqrt{1 + \beta \frac{R_d}{R_{eq}}}\right)$$

$$= 2\left[\frac{R_{eq}}{R_d} + \sqrt{\left(\frac{R_{eq}}{R_d}\right)^2 + \beta\frac{R_{eq}}{R_d}}\right]$$

$$= \eta_{OPT} - 1 \qquad \text{from (4.44).}$$

So that $$\eta_{OPT} = 1 + 4\pi CB_0R_{eq} \quad . \quad . \quad . \quad (4.67)$$

Comparison between equations (4.67) and (4.64) is instructive. It shows that for a given band-width, a lower noise factor can always be achieved at low frequencies than at frequencies where the valve has an appreciable input conductance $1/R_d$, even if matching for optimum noise factor is not possible at the lower frequency. This is an important consideration in the choice of the frequency of the I.F. amplifier for receivers with a wide band-width.

Broad Band Circuits. It may occur that the band-width under optimum noise factor conditions, given by equation (4.66), is insufficient for the application envisaged. Two main types of system need broad band circuits.

(1) Systems in which the intelligence spectrum is wide, particularly television, pulse modulation and radar systems.

(2) Systems in which the exact carrier frequency may be unknown or subject to wide tolerances.

Equation (4.61) shows that an increase in band-width may be achieved by a decrease of the circuit capacity C, by a decrease of the valve input resistance R_d, or by a decrease of the transformer step-up ratio *m*, assuming that the source impedance R_a is fixed. The circuit capacity cannot be decreased beyond a certain value representing the valve input capacity and the equivalent shunt capacity of the tuning device; there are therefore limitations for any increase in band-width along this line.

A decrease of the valve input impedance R_d may easily be accomplished by the addition of an external shunt load. This leads, however, to a serious deterioration of the noise factor, as may be seen from equation (4.42); it is physically obvious that the shunt load attenuates the signal before it reaches the grid, whereas the noise generated in the valve is still at its original level. This method of increasing the band-width must therefore be ruled out as objectionable.

It will be shown later in Section 4.3.2, that by the use of a cathode input-amplifier, an effective extra shunt load can be thrown across the tuned circuit which will materially increase band-width without deteriorating the noise factor.

Finally, the band-width may be increased by lowering the transformer step-up ratio *m*, i.e. coupling the input more tightly to the circuit than the optimum noise factor condition (4.43) indicates. A considerable increase in band-width is possible by this method with only a slight deterioration of noise factor, as a quantitative examination of equation (4.42) will show.

High-selectivity Circuits. In extensive multi-channel communication systems, it is usually essential to accommodate a large number of channels in a restricted frequency range. The transmitter frequencies are then usually stabilized by direct or indirect crystal control methods, and are held to narrow tolerances. The spectrum required by telephone or automatic telegraphy channels with amplitude-modulation or low-deviation frequency-modulation is relatively small, and it thus becomes possible to space adjacent communication channels very closely.

The selectivity required in the receiver to reject undesired signals under these conditions is of course mainly achieved in the I.F. amplifier. Two kinds of interfering signal must, however, be coped with by adequate selectivity in the R.F. amplifier. The first of these is a signal on the *image frequency* f_i ($f_i = 2f_h - f_s$, where f_s is the signal frequency, f_h the local oscillator frequency). A signal entering the receiver on this image frequency will have an output component at the I.F. after passing through the frequency-converter stage, and must therefore be adequately attenuated before reaching the converter. This is facilitated by the use of the highest possible intermediate frequency f_0, since $|f_i - f_s| = 2f_0$.

Cross-modulation. The second kind of interference which necessitates high selectivity in the R.F. circuits, particularly in the input stage itself, will only be experienced when the level of the interfering signal is extremely high, i.e. of the order of one volt or so at the grid of the first valve. This is most likely in marine applications, where a strong local transmitter may be in operation on a neighbouring channel during reception. Under these conditions, the non-linear characteristic of the first valve can transfer some of the modulation from the interfering carrier on to the desired carrier—a phenomenon known as *cross-modulation.*[12], [13]

I.F. selectivity will contribute nothing to the elimination of this interference, since it is modulated on to the desired carrier itself. The only real remedy is to attenuate the interfering signal as much as possible before it reaches the first valve, i.e. a very high-selectivity input circuit. A rather more crude method is to attenuate *all* signals before reaching the input valve. This will of course reduce the sensitivity of the receiver, but reduces cross-modulation much more rapidly since the cross-modulation factor is proportional to the square of the interfering signal voltage at the first grid. Thus a uniform attenuation of 5 db will reduce cross-modulation by 10 db. Other methods of reducing cross-modulation have been proposed,[13] but are unsuitable for U.H.F. It can be shown that cross-modulation in the strict sense defined above does not occur with frequency-modulated signals, so that it can be disregarded in F.M. systems.

If the band-width for optimum noise factor of the circuit of Figure 47A, given by equation (4.66), is too wide in view of the foregoing considerations, equation (4.61) shows that it can be reduced by increasing the circuit capacity C, or the transformer step-up ratio m.

An increase in C may be permissible, but must be accompanied by a decrease of the equivalent circuit inductance L to maintain the circuit at the same frequency as before. This lowering of the L/C ratio will, however, decrease the intrinsic dynamic impedance of the circuit, which has hitherto been regarded as sufficiently high to be disregarded in comparison with the load R_d thrown across the circuit by the valve, and an additional loss may therefore be introduced. Quite apart from this, there are natural limits to the process of reducing the equivalent inductance L, which apply both for lumped circuits and resonant lines or cavities, and it may become impossible to maintain the circuit resonant at a sufficiently high frequency. For circuits required to tune over a wide frequency range it is particularly desirable to keep C low. If the transformer step-up ratio m is increased, i.e. the source is tapped down further, and is therefore more loosely coupled to the circuit, the loading of the source impedance across the circuit will be reduced, and a more sharply selective circuit will result. But this will be accompanied by a loss in signal transfer, and a consequent loss in sensitivity (equation (4.42)); and even if the source were tapped down to the absolute limit ($m \to \infty$), the resultant band-width could only be reduced to a minimum value $B = 1/2\pi CR_d$ given by the damping of the circuit by the valve.

In order to achieve really high selectivities, it is therefore essential to reduce the damping effect of the valve as well as that of the source, by tapping both the source and the valve itself down on the circuit. Consider therefore the circuit of Figure 50A, in which the valve is tapped down on the circuit in the ratio $1 : n$, and the source is tapped down on the circuit in the ratio $1 : m\,n$, so that the step-up ratio between source and valve is again m as before.

Under these conditions the intrinsic dynamic circuit impedance R_u may no longer be neglected, as is evident from Figure 50B. Previously, the valve input impedance R_d was directly shunted across R_u, and since we may always assume that R_u is much higher than R_d, the combination of the two impedances in shunt was practically R_d. Now, the transferred valve impedance seen by the circuit is n^2R_d, where n may be large, and unless $R_u \gg n^2R_d$, we must take R_u into account.

In place of equation (4.61) we now have for the band-width

$$B = \frac{1}{2\pi C}\left(\frac{1}{m^2 n^2 R_a} + \frac{1}{R_u} + \frac{1}{n^2 R_d}\right) \quad . \quad . \quad (4.68)$$

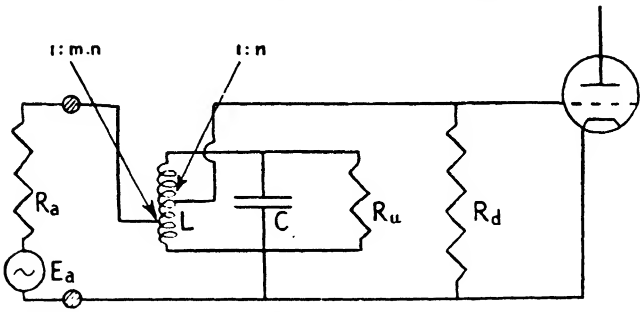


Fig. 50A. Receiver input stage arranged to give high selectivity

The noise factor analysis for this circuit may be carried out exactly as for the circuit of Figure 47A, using the equivalent circuit referred to the *valve* in Figure 50C, and bearing in mind that R_u/n^2, the circuit impedance referred to the valve, is itself a thermal noise source. The detailed analysis will be omitted here, but leads to the

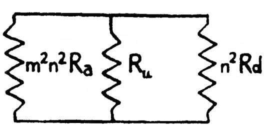


Fig. 50B. Equivalent circuit of the stage of Fig. 50A, omitting generators, referred to the circuit

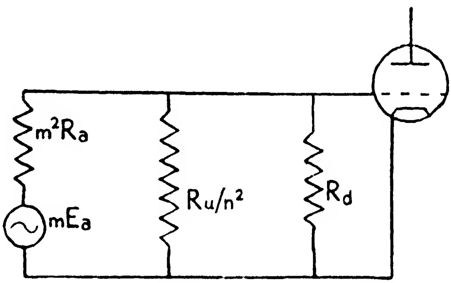


Fig. 50C. Complete circuit equivalent of the stage shown in Fig. 50A, referred to the valve

conclusion that the best noise factor for a given overall band-width is obtained for a step-up ratio m between *source* and *valve* which still has the same value m_0 as that given in equation (4.43). The step-up ratio n from the valve to the circuit is then determined by the selectivity requirements from equation (4.68). It is evident that if the valve is tapped sufficiently far down on the circuit (n being large), the band-width may be narrowed right down until it is comparable with that of the unloaded circuit.

Figure 50C shows that the effect of the circuit impedance reflected across the valve is an insertion loss equivalent to a resistance R_u/n^2 across R_d. Since m has the same value as previously, the signal reduction at the grid is simply that due to the insertion of this extra shunt impedance R_u/n^2 across the circuit, which can be deduced from the circuit; and the deterioration of the noise factor may be put approximately equal to the insertion loss caused by R_u/n^2. In order to minimize this loss, it is essential to use circuits of the highest possible intrinsic dynamic impedance R_u (i.e. the highest possible unloaded Q), since it may then be possible to tap both valve and source well down on the circuit and thus to achieve high selectivity,

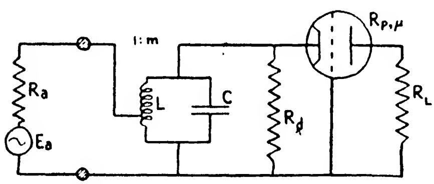


Fig. 51A. Receiver input stage employing grounded-grid valve

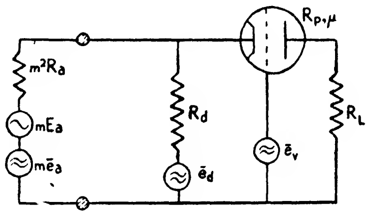


Fig. 51B. Stage shown in Fig. 51A showing the noise generators

and yet to keep the transferred shunt impedance R_u/n^2 across the valve high compared with the valve input impedance R_d. So long as R_u/n^2 is greater than $4R_d$, the noise factor deterioration caused by tapping down the valve will be less than 1 db at all times.

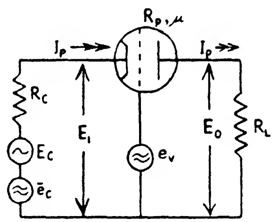


Fig. 51C. The equivalent of Fig. 51B including R_d in the cathode resistance R_c

4.3.2. *The Cathode-input Amplifier.* It has been shown that the noise factor of the triode is superior to that of the pentode for U.H.F. signal-frequency amplifiers. But the triode when used as a grid-input amplifier needs careful neutralization of the feed-back capacity between input and output circuit, which in this case is the grid-anode capacity of the valve. Neutralized circuits have been successfully used, but the adjustment is very critical even for a fixed frequency circuit, and is impossible for a circuit with a wide tuning range. For this reason, it is now universal practice to use triodes in the grounded-grid condition, in which the signal is fed into the cathode; disc-seal or "light-house" construction allows the planar grid-structure to be made an integral part of the external screen between input and output circuits, and the anode-cathode capacity can be made effectively so small that no neutralization is required.

The required modifications of the results of Section 4.3.1 for cathode-input amplifiers will now be developed.

4.3.2.1. *Impedance and Gain Relations.* A schematic circuit of a cathode-input stage is shown in Figure 51A, with the source tapped down in the circuit as in Figure 47A, giving a step-up ratio m from source to circuit. The valve in the first instance is applied across the whole circuit. R_d as before is the high-frequency input impedance across the grid-cathode terminals, which may be neglected at low frequencies. It will be seen presently that an additional input load appears in the grid-cathode circuit by the feed-back effect of the cathode current. It is therefore best, in the interests of clarity, to regard the high-frequency loading R_d as an *external* shunt resistance across the cathode-grid terminals, and to re-draw the equivalent circuit of Figure 51B as shown in Figure 51C.

The external circuit is now represented by an R.M.S. signal E.M.F. E_c acting in series with an impedance R_c. Their values are

$$R_c = \frac{m^2R_aR_d}{m^2R_a + R_d} \text{ as before} \quad . \quad . \quad . \quad (4.69)$$

$$E_c = mE_a\frac{R_d}{m^2R_a + R_d} \quad . \quad . \quad . \quad (4.70)$$

If the plate signal current is I_p, the actual R.M.S. signal voltage at the cathode will be, from Figure 51C,

$$E_1 = E_c - R_cI_p \quad . \quad . \quad . \quad (4.71)$$

The internal plate resistance of the valve is R_p, the final load resistance R_L.

The R.M.S. plate signal current in the direction of the arrows indicated in Figure 51C will be given by

$$I_p = gE_1 + \frac{E_c - I_p(R_c + R_L)}{R_p}$$

$$= g(E_c - R_cI_p) + \frac{E_c - I_p(R_c + R_L)}{R_p}$$

Hence $\quad I_pR_p = \mu(E_c - R_cI_p) + E_c - I_p(R_c + R_L)$

and

$$I_p = \frac{(\mu + 1)E_c}{R_p + R_L + (\mu + 1)R_c} \quad . \quad . \quad . \quad (4.72)$$

Equation (4.72) may be written

$$I_p = \frac{E_c}{R_c + R_f} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.73)$$

where

$$R_f = \frac{R_p + R_L}{\mu + 1} \approx \frac{R_p + R_L}{\mu} = \frac{1}{g} + \frac{R_L}{\mu} \quad . \quad (4.74)$$

Equation (4.73) shows that R_f is the *feed-back input impedance* of the valve seen by the source, which is caused by the flow of the plate current I_p through the source. This feed-back input impedance R_f is independent of, and additional to, the high-frequency input impedance R_d which we are now regarding as an external shunt impedance; and the feed-back input impedance R_f will arise even at low frequencies where R_d may be high enough to be neglected. The total input load of the cathode-input stage appearing across the tuned circuit is therefore given by R_d and R_f in shunt, as shown in Figure 51D. Equation (4.74) shows that the feed-back input impedance depends to some extent on the final output load impedance R_L; but if $R_L \ll R_p$, R_f is approximately equal to $1/g$, and is therefore very low for a high-slope valve.

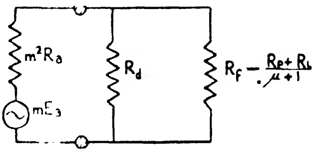


Fig. 51D. Total input load of a grounded-grid stage

Alternatively, equation (4.72) may be written

$$I_p = \frac{(\mu + 1)E_c}{R_0 + R_L} \quad . \quad . \quad . \quad (4.75)$$

where

$$R_0 = R_p + \overline{\mu + 1}\, R_c \quad . \quad . \quad . \quad (4.76)$$

so that R_0 is the *output impedance* of the valve, as seen by the load R_L; and equation (4.76) shows that the output impedance is a function of the source impedance as well as of the plate impedance.

Gain. The power gain of the stage is again defined as $G = P_0/P_i$ where P_0 is the available output power, P_i the available input power. The available output power P_0 follows from equation (4.75)

$$P_0 = \frac{(\overline{\mu + 1}\, E_c)^2}{4R_0}$$

$$= \frac{\{(\mu + 1)\, E_c\}^2}{4\{R_p + (\mu + 1)\, R_c\}} \quad . \quad . \quad . \quad (4.77)$$

For the available input power we have, as previously,

$$P_i = \frac{(mE_a)^2}{4m^2R_a} = \frac{E_a^2}{4R_a} \quad . \quad . \quad . \quad . \quad (4.49)$$

With the aid of equations (4.49), (4.69), (4.70), and (4.77), the stage gain can be evaluated in terms of R_a, m, R_d, g and R_p.

The maximum stage gain. This is of particular interest. It will be attained when the source is matched to the total input impedance

(R_d and R_f in shunt), and when the load is matched to the output impedance. These two conditions are

$$m^2R_a = \frac{R_dR_f}{R_d + R_f}$$

$$= \frac{R_d(R_p + R_L)}{R_p + R_L + \overline{\mu + 1}\,R_d}$$

$$= R_d\frac{R_p + R_L}{M^2R_p + R_L} \quad . \quad . \quad . \quad . \quad (4.78)$$

where
$$M^2 = 1 + \overline{\mu + 1}\,\frac{R_d}{R_p} = 1 + \left(g + \frac{1}{R_p}\right)R_d \quad . \quad (4.79)$$

and
$$R_L = R_0 = R_p + \overline{\mu + 1}\,R_c$$

$$= R_p + \overline{\mu + 1}\,R_d\frac{m^2R_a}{R_d + m^2R_a}$$

$$= R_p\frac{R_d + M^2m^2R_a}{R_d + m^2R_a} \quad . \quad . \quad . \quad . \quad (4.80)$$

Equations (4.78) and (4.80) can be solved for R_L and m^2R_a, giving for maximum stage gain

$R_L = MR_p$, determining the load impedance (4.81)

$m^2R_a = R_d/M$, determining the step-up ratio m . (4.82)

For this value of m,

$$R_c = \frac{m^2R_aR_d}{m^2R_a + R_d} = \frac{R_d}{M + 1} \quad . \quad . \quad (4.83)$$

and
$$\frac{E_c}{mE_a} = \frac{R_c}{m^2R_a} = \frac{M}{M + 1} \quad . \quad . \quad (4.84)$$

The available output power is now

$$P_0 = \frac{\{(\mu + 1)E_c\}^2}{4R_0} = \frac{(\mu + 1)^2\left(\frac{M}{M + 1}\right)^2(mE_a)^2}{4MR_p}$$

and the available input power is

$$P_i = \frac{(mE_a)^2}{4m^2R_a} = \frac{M(mE_a)^2}{4R_d}$$

so that the maximum stage gain is

$$G_{max} = \frac{P_0}{P_i} = \left(\frac{\mu + 1}{M + 1}\right)^2 \cdot \frac{R_d}{R_p} \quad . \quad . \quad (4.85)$$

Comparing this with the maximum stage gain attainable with the grid input stage (equation (4.58)), we obtain for the ratio

$$\frac{G_{max\,(\text{cathode input})}}{G_{max\,(\text{grid input})}} = \left(\frac{\mu + 1}{\mu}\right)^2 \times \frac{4}{(M + 1)^2}$$

$$\approx \frac{4}{(M + 1)^2} \quad . \quad . \quad . \quad (4.86)$$

Since $M \approx (1 + gR_d)^{\frac{1}{2}} > 1$, the stage gain possible with a cathode-input stage is therefore lower than that for the same valve used as a grid-input amplifier. As R_d decreases with increasing frequency, however, $M \to 1$, and at the higher frequencies the difference between the theoretical maximum gains becomes insignificant.

4.3.2.2. *Determination of Optimum Noise Factor.* The noise sources for the cathode-input circuit of Figure 51A are indicated in Figures 51B and 51C. As before, we have

$$\overline{e_a^2} = 4kT_0R_aB \quad . \quad . \quad . \quad . \quad (4.33)$$

$$\overline{e_d^2} = 4k\beta T_0R_dB \quad . \quad . \quad . \quad . \quad (4.34)$$

$$\overline{e_v^2} = 4kT_0R_{eq}B \quad . \quad . \quad . \quad . \quad (4.35)$$

and we again put $\quad R_d = p^2R_a \quad . \quad . \quad . \quad . \quad (4.36)$

But note that p is now no longer the step-up ratio for maximum power transfer, which may be denoted by q and follows from equation (4.82)

$$q = p/\sqrt{M}$$

The mean square noise E.M.F. $\overline{e_c^2}$ in Figure 51C is derived from those in Figure 51B by the relation

$$\overline{e_c^2} = m^2\overline{e_a^2}\left(\frac{R_d}{m^2R_a + R_d}\right)^2 + \overline{e_d^2}\left(\frac{m^2R_a}{m^2R_a + R_d}\right)^2$$

$$= m^2\overline{e_a^2}\left(\frac{p^2}{m^2 + p^2}\right)^2 + \overline{e_d^2}\left(\frac{m^2}{m^2 + p^2}\right)^2 \quad . \quad . \quad (4.87)$$

From equation (4.72), this noise E.M.F. $\overline{e_c}$ will give rise to a noise current component in the plate circuit of mean square value

$$\overline{i_c^2} = \frac{(\mu + 1)^2\overline{e_c^2}}{\{R_p + R_L + (\mu + 1)R_c\}^2}$$

Similarly, the shot noise E.M.F. $\overline{e_v}$ in the grid will, from conventional circuit analysis, give rise to a noise current component in the plate circuit of mean square value

$$\overline{i_v^2} = \frac{\mu^2\overline{e_v^2}}{\{R_p + R_L + (\mu + 1)R_c\}^2}$$

If we neglect the thermal noise of the load impedance R_L, as before, we have for the output signal/noise power ratio

$$\frac{I_p^2}{\overline{i_p^2}} = \frac{I_p^2}{\overline{i_c^2} + \overline{i_v^2}} = \frac{(\mu + 1)^2 E_c^2}{(\mu + 1)^2\overline{e_c^2} + \mu^2\overline{e_v^2}}$$

$$= \frac{E_c^2}{\overline{e_c^2} + \left(\frac{\mu}{\mu + 1}\right)^2 \overline{e_v^2}} \quad . \quad . \quad (4.88)$$

The noise factor now becomes

$$\eta = \frac{E_a^2}{\overline{e_a^2}} \Big/ \frac{I_p^2}{\overline{i_p^2}} = \frac{E_a^2}{E_c^2} \times \frac{\overline{e_c^2} + \left(\frac{\mu}{\mu + 1}\right)^2 \overline{e_v^2}}{\overline{e_a^2}}$$

which gives, by combining equations (4.33), (4.34), (4.35), (4.36), (4.70) and (4.87), and simplifying:

$$\eta = 1 + \beta\frac{m^2}{p^2} + \left(\frac{\mu}{\mu + 1}\right)^2 \left(\frac{p}{m} + \frac{m}{p}\right)^2 \frac{R_{eq}}{R_a} \qquad (4.89)$$

If $\mu \gg 1$, so that

$$\left(\frac{\mu}{\mu + 1}\right)^2 \approx 1,$$

the expression for the noise factor of the cathode-input stage is therefore completely identical with the corresponding expression for the grid-input stage, if the transformer step-up ratio m is the same. This result might be regarded as physically evident, since the feedback input impedance R_f in Figure 51D attenuates the signal current and the noise currents approximately equally, so that R_f can be regarded as a *noiseless* resistance across the circuit.

In particular, it leads to the conclusion that the step-up ratio m_0 for optimum noise factor is still given by

$$\frac{m_0^4}{p^4} = \frac{R_{eq}}{R_{eq} + \beta R_a} \quad . \quad . \quad . \quad (4.43)$$

and that for this step-up ratio we still have, for the optimum noise factor itself

$$\eta_{OPT} = 1 + 2\left[\frac{R_{eq}}{R_a} + \sqrt{\left(\frac{R_{eq}}{R_a}\right)^2 + \beta\frac{R_{eq}}{R_a}}\right] \quad . \quad (4.44)$$

and, finally, that the curve plotted in Figure 48 for the optimum noise factor as a function of frequency is still valid for the cathode-input stage.

If we express the optimum step-up ratio m_0 in terms of the new maximum power transfer ratio q given by equation (4.82), we have

$$\frac{m_0^4}{q^4} = M^2\frac{m_0^4}{p^4} = M^2\frac{R_{eq}}{R_{eq} + \beta R_a} \quad . \quad . \quad (4.90)$$

and since $M > 1$, the optimum step-up ratio for the cathode-input circuit may be *higher* than the maximum power transfer ratio q; i.e. whereas for the grid-input amplifier the optimum coupling was tighter than that for maximum power transfer, the optimum coupling for the cathode-input amplifier may be *more loose* than that for maximum power transfer. Note that the actual step-up for optimum noise factor is the same in the two cases; it is the step-up for maximum power transfer which differs in the two cases.

4.3.2.3. *Band-width Considerations.* The band-width of the cathode-input circuit between half-power points is given, as before, by $B = 1/2\pi CR_t$, where R_t is the total shunt impedance across the tuned circuit. From Figure 51D we see that this consists of the transferred source impedance m^2R_a, the high-frequency input impedance R_d, and the feed-back input impedance R_f in shunt. We therefore get, for the circuit of Figure 51, a band-width

$$B = \frac{1}{2\pi C}\left(\frac{1}{m^2R_a} + \frac{1}{R_d} + \frac{1}{R_f}\right)$$

$$= \frac{1}{2\pi C}\left(\frac{1}{m^2R_a} + \frac{1}{R_d} + \frac{\mu + 1}{R_p + R_L}\right) \quad . \quad (4.91)$$

It is evident from (4.91) and from Figure 51D that the additional feed-back input impedance R_f shunted across the circuit increases the band-width considerably beyond that of the equivalent grid-input stage, and since R_f may be regarded as noiseless, the band-width is increased without any deterioration of the noise factor.

At first sight it would appear that this effect makes the cathode-input amplifier eminently suitable for use in cascade in a multi-stage wideband I.F. amplifier. It is found, however, that when using more than one cathode-input stage in cascade, much of the advantage is lost because of the low input impedance of the second and later stages. In these circumstances, it is difficult or impossible to attain the maximum gain for the first stage, and noise from the second and later stages makes an appreciable contribution. In practice, a combination of grid-input and cathode-input stages has been used for wideband I.F. amplifiers with advantage.[7,11]

High-selectivity Circuits. If high selectivity is required in a cathode-input stage, the same considerations apply as in Section 4.3.1.3. It becomes essential to tap down the valve, as well as the source, on the tuned circuit; and the best noise factor will again be attained if this is done in such a manner that the step-up ratio between *source* and *valve* has the same value m_0 determined for optimum noise factor in the previous case. As before, the intrinsic dynamic impedance of the tuned circuit may have to be taken into account, and will cause an insertion loss in the input circuit unless sufficiently high.

The unloaded Q of the tuned circuit should therefore have the highest possible value to avoid a loss in sensitivity. Since the analysis is exactly parallel to that of the grid-input circuit previously discussed, it will be omitted here.

4.4. **U.H.F. Mixer and Local Oscillator Circuits.**[16] The mixer circuit transposes the R.F. signal down to the intermediate frequency by the injection of a locally generated heterodyne frequency and the action of the non-linear transconductance characteristic of the mixer device itself.

The design of the mixer stage needs careful attention in receivers where the mixer is not preceded by a signal-frequency amplifier, which is the normal case for receivers working above 1000 Mc/s, and is very often the case for receivers at much lower frequencies if economy and simplicity of design and operation are important. The sensitivity of the receiver is then determined by the mixer circuit and the input stages of the I.F. amplifier. The R.F. selectivity of such a receiver will generally be poor, and a reasonable image rejection can only be achieved by using a high intermediate frequency.

The conversion gain G of the mixer stage is defined by the relation

$$G_c = P_0/P_i \qquad . \quad . \quad . \quad . \quad (4.92)$$

where P_0 is the available output power at the I.F., and P_i is the available input power at the signal frequency. The most important requirements for a mixer are therefore low noise factor, high conversion gain, and low power absorption from the local oscillator.

Multi-electrode and pentode mixer stages are hardly ever used for U.H.F. applications, as their efficiency is very low, and their noise factor poor. Triode mixers are used to some extent up to a few hundred Mc/s; but, in general, diode or crystal mixers are in universal use for frequencies above 300 Mc/s. The technique, common at low frequencies, of combining local oscillator and mixer in a single converter valve is impracticable for U.H.F. A brief account of triode, diode, and crystal mixers will be given.

4.4.1. *Triode Mixers.* It is assumed that the reader is familiar with the conventional theory of the triode mixer[12] and particularly with the concept of the conversion transconductance g_c, which is the ratio of the I.F. plate current (with a short-circuited load) to the signal-frequency grid-voltage. The conversion transconductance varies with the input voltage from the local oscillator; and if the peak local oscillator voltage is always equal to the standing negative bias voltage on the grid, g_c increases with increasing local oscillator voltage to a maximum, and will then decrease very slowly as the local oscillator voltage increases further.

In order to derive the equivalent noise resistance of the mixer triode, we must make the necessary modifications to equations (4.11) and (4.14) in Section 4.2.4.2. Since the mutual conductance g of the valve varies rhythmically during the local oscillator cycle, our expression for the mean square (I.F.) noise current, previously given in equation (4.11), now becomes

$$\overline{i_a^2} = 0{\cdot}64 \times 4kT_{cath}\bar{g}B \qquad . \quad . \ (4.93)$$

where $\bar{g}$ is the *average* mutual conductance during the local oscillator cycle.

The equivalent noise resistance R_{eq} is again defined by the noise voltage referred to the grid.

$$\overline{e_v^2} = 4kT_0R_{eq}B \qquad . \quad . \quad . \ (4.16)$$

and we now have

$$\overline{i_a^2} = g_c^2\,\overline{e_v^2}, \qquad . \quad . \quad . \quad . \ (4.94)$$

where g_c is the conversion conductance.

Hence
$$R_{eq} = \frac{\overline{i_a^2}}{g_c^2 \,.\, 4kT_0B} \qquad . \quad . \quad . \ (4.95)$$

If we take $T_{cath} = 1150°$ K, conductances in mA/V, the equivalent noise resistance R_{eq} in kilo-ohms becomes, from (4.93) and (4.95)

$$R_{eq} = \frac{2{\cdot}5\bar{g}}{g_c^2} \qquad . \quad . \quad . \quad . \ (4.96)$$

If the maximum mutual conductance during the L.O. cycle is g_{max}, we find that for operation at the fundamental of the local oscillator frequency, the optimum value for g_c is approximately $g_c = 0{\cdot}28\ g_{max}$; and if we put for the mean slope $\bar{g} = 0{\cdot}5\ g_{max}$ approximately, this gives

$$R_{eq} = \frac{16}{g_{max}} \text{ approximately} \qquad . \quad . \ (4.97)$$

Comparing this result with equation (4.15A), we see that the equivalent noise resistance for a mixer is substantially higher than for the same valve used as an amplifier, giving a much poorer noise-factor; and since the conversion gain of the mixer is much lower than the amplifier gain, the noise from the I.F. input stage will usually have to be taken into account with the aid of equation (4.28) in computing the overall noise factor of the receiver, if the mixer is the first stage. The evaluation of the noise factor proceeds as in Section 4.3, using the value for R_{eq} given by (4.96) or (4.97).

The conversion gain itself is calculated exactly as in Section 4.3.1.2, by substituting the conversion conductance g_c for the mutual conductance g in all the relevant formulae. This gives for the maximum possible conversion gain, by comparison with equation (4.58)

$$G_{c\,(max)} = \frac{g_c{}^2 R_p R_d}{4} \quad . \quad . \quad . \quad (4.98)$$

where R_p is the plate impedance, R_d the high frequency input impedance of the valve.

4.4.2 *Diode Mixers.* At frequencies above 300 Mc/s, the conversion gain of the triode mixer becomes very poor, and the low

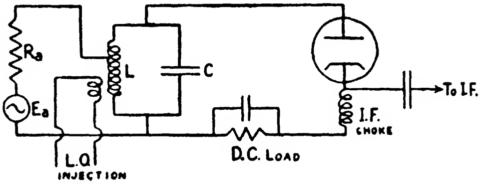


Fig. 52A. Diagram of a diode mixer stage

input impedance makes considerable demands on the power required from the local oscillator if optimum conversion efficiency is to be achieved, since it is usually necessary to swing the grid by several volts. It is therefore better to use a diode as the mixer, although the conversion gain with this will necessarily be less than unity.

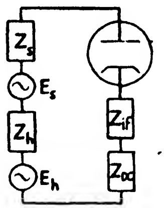


Fig. 52B. Circuit equivalent of the stage shown in Fig. 52A

The schematic of a diode mixer stage and its equivalent circuit is given in Figures 52A and 52B. In Figure 52B, E_s and Z_s denote the transferred signal E.M.F. and source impedance in the diode circuit, E_h and Z_h the transferred local oscillator E.M.F. and source impedance, Z_{if} the I.F. load impedance, and Z_{dc} the D.C. resistance. The analysis of the circuit is made complex by the fact that the input and output currents flow in the same circuit loop, and interact on each other owing to the non-linear diode characteristic. For a full treatment, the reader is referred to the work of James and Houldin.[(14)] The main results only will be stated here.

If the diode current characteristic is $i = f(v)$, then the instantaneous conductance is

$$g = \frac{di}{dv} = f'(v)$$

Now if the local oscillator voltage E_h is periodic with angular frequency ω_h, and is large compared with the signal voltage E_s, the conductance may be assumed to be unaffected to the first order by the signal voltage, but to vary rhythmically with the local oscillator voltage, so that it may be represented by a Fourier expansion

$$g = g_0 + g_1 \cos \omega_h t + g_2 \cos 2\omega_h t + \ldots \quad (4.99)$$

where g_0, g_1, g_2, etc., depend on the amplitude of the local oscillator signal, and also on the D.C. bias voltage developed in Z_{dc}. g_0 is seen to be the average conductance of the diode over the local oscillator cycle.

The conversion conductance g_c is the ratio of I.F. output current to signal input voltage, if both source impedance and I.F. load impedance are shunted out. If a small signal $E_s \cos \omega_s t$ is applied to the diode, the instantaneous current will be

$$\begin{aligned} i &= gE_s \cos \omega_s t \\ &= E_s \cos \omega_s t[g_0 + g_1 \cos \omega_h t + g_2 \cos 2\omega_h t + \ldots] \end{aligned}$$

If the I.F. is derived from the local oscillator fundamental, this gives for the amplitude of the current component at the I.F.

$$i_0 = \frac{g_1}{2} E_s$$

Therefore

$$g_c = \frac{i_0}{E_s} = \frac{g_1}{2} \quad (4.100)$$

Similarly, if the n^{th} harmonic of the oscillator is used, we have

$$g_c = \frac{g_n}{2} \quad (4.101)$$

The conversion conductance g_c is always smaller than the average conductance g_0. If the L.O. current pulses through the diode are very short compared with the L.O. period, $g_c/g_0 \to 1$ in the limit as the pulse length approaches zero; but in this case both g_c and g_0 tend to zero.

The conversion conductance for higher harmonics is smaller than for operation at L.O. fundamental, but this again tends to become very nearly equal to g_0 for very short L.O. current pulses.

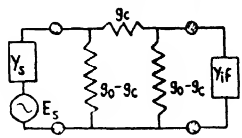


Fig. 52c. Linear circuit equivalent of the stage shown in Fig. 52A as an admittance diagram

Although the diode circuit is non-linear, it may be replaced by a linear equivalent circuit, which is valid for any particular local oscillator injection determining the Fourier series (4.99) for the diode conductance g, so long as the signal voltage is small compared with the local oscillator voltage. The equivalent circuit is shown in an admittance diagram in Figure 52c.

The diode itself appears as a π — section, in which the horizontal arm is the conversion conductance g_c, the vertical arms are equal to $g_0 - g_c$. Since g_0 is always greater than g_c, all the arms are positive conductances. From the equivalent circuit in Figure 52C, the signal-frequency input impedance, the I.F. output impedance, the conversion gain and the optimum matching conditions may all be derived; and we may proceed exactly as if all currents in the network were at the same frequency, although in fact a frequency transposition takes place.

If the local oscillator coupling is sufficiently loose to neglect its reflected impedance, Y_s is the signal-frequency admittance of the source, and Y_{if} is the I.F. admittance of the load, then the input admittance of the diode to the signal is

$$Y_{in} = g_0 - \frac{g_c^2}{g_0 + Y_{if}} \quad . \quad . \quad . \quad (4.102)$$

and the output admittance at the intermediate frequency is

$$Y_{out} = g_0 - \frac{g_c^2}{g_0 + Y_s} \quad . \quad . \quad (4.103)$$

The conversion gain G_c will be a maximum if $Y_s = Y_{in}$, and $Y_{out} = Y_{if}$.

It is then found that $Y_s = Y_{in} = Y_{out} = Y_{if} = (g_0^2 - g_c^2)^{\frac{1}{2}}$, and the conversion gain is now

$$G_{c(max)} = \frac{g_c^2}{(g_0^2 + \sqrt{g_0^2 - g_c^2})^2} \quad . \quad . \quad (4.104)$$

This is clearly always less than unity, and can only approach unity if $g_c \rightarrow g_0$, which leads to infinite impedances throughout and is therefore not a practical case.

In practice, there is always a conversion loss which is rarely less than 6 db. The conversion loss, if expressed as a power ratio, is $L_c = 1/G_c$.

The equivalent diagram of Figure 52C shows no circuit losses. If these are to be taken into account, they can be represented by another shunt conductance across the input terminals of the diode; and the equivalent circuit can again be used for the derivation of the impedance relationships and the conversion gain.

Noise Factor of Diode Mixer. The noise output of the space-charge limited diode, when operating as a U.H.F. mixer, is generally higher than that predicted by the theory of Section 4.2.4. A number of factors contribute to this effect, and they are not yet sufficiently clearly analysed to make a quantitative treatment feasible within the scope of this book. We therefore characterize the diode empirically by its noise temperature ratio β, defined as the ratio of the available

I.F. noise power of the diode to the available noise power from a resistance at room temperature, equal to the output impedance of the diode. β is measured with the normal excitation from the local oscillator on the diode.

If the noise factor of the I.F. amplifier following the mixer is η_2, the diode must supply signal power

$$(\eta_2 + \beta - 1)kT_0B$$

to the I.F. amplifier for unity signal/noise ratio at the output. If the conversion loss in the diode is L_c, the available signal power at the input must be

$$L_c(\eta_2 + \beta - 1)kT_0B$$

The noise factor of the receiver is therefore

$$\eta = L_c(\eta_2 + \beta - 1) \quad \ldots \quad (4.105)$$

For a typical diode with a noise temperature ratio $\beta = 2{\cdot}5$, a conversion loss of 6 db ($L_c = 4$), and an I.F. amplifier of noise factor 5 db ($\eta_2 = 3{\cdot}2$), we get

$$\eta = 18{\cdot}6(\sim 12{\cdot}7 \text{ db})$$

This is about the limiting sensitivity that can be achieved at present with diode mixers.

4.4.3. *Crystal Mixers.* Within the last few years, the silicon-tungsten crystal has almost completely replaced all other forms of mixers at microwave frequencies. This surprising resurrection of one of the oldest devices used in radio has been made possible by intensive research during the War, which has produced a reliable, compact, robust component that has little in common with the old "cat's-whisker," except the basic principle of rectification at the barrier layer between the silicon semi-conductor and the tungsten contact wire.[15] It is possible to set the contact in manufacture at a point of sufficient sensitivity, so that no adjustments are necessary in operation, and the crystal unit is mounted in a sealed cartridge. Nevertheless, the main problem to-day is still the uniformity of performance of the manufactured article.

Compared with the diode, the crystal has several advantages. Its shunt capacity can be made much smaller than that of the diode; there are no transit-time effects to set an upper limit to the operating frequency; it is physically smaller and can therefore be more easily built in as an integral part of the mixer circuit; it needs no filament supply, and thus does away with the problem of decoupling the supply leads; and lastly its more sharply-curved characteristic allows operation with smaller excitation power from the local oscillator. The main disadvantage is its liability to burn out under heavy overloads, such as occur in radar sets with common T/R aerial working. The overloads likely to be met in communication applications should not damage the crystal. Crystals using germanium instead of silicon are more robust in this respect, but at present their noise factor is inferior.

The circuit analysis given for the diode mixer in Section 4.4.2 is valid for the crystal mixer also. The conversion conductances for fundamental, second, and higher harmonic mixing are derived from the Fourier expansion of the crystal conductance g; the impedances and the conversion loss may be deduced from the equivalent circuit of Figure 52c. The conversion loss of the crystal mixer decreases with increasing excitation down to a minimum of about 6 db; the excitation is usually measured by the D.C. current in the crystal circuit. For crystal currents between 0·5 and 1·0 mA, the conversion loss usually is between 6 and 10 db for operation at oscillator fundamental. For second harmonic operation, the conversion loss is at least 3 db greater in practice. Unlike the diode, the crystal has a finite back-to-front resistance ratio, which may be as low as 10 : 1, although it is usually a good deal higher. This makes the crystal less efficient as a mixer for operation at the higher oscillator harmonics.

The noise temperature ratio β of the crystal is again used to specify the noise factor of the mixer circuit, as in the case of the diode. It is found experimentally that β increases very nearly linearly with crystal current. For zero crystal current, $\beta = 1$; for crystal currents of 1 mA, $\beta \sim 3$, but this varies considerably from one crystal to another. The noise factor of the receiver is again given by

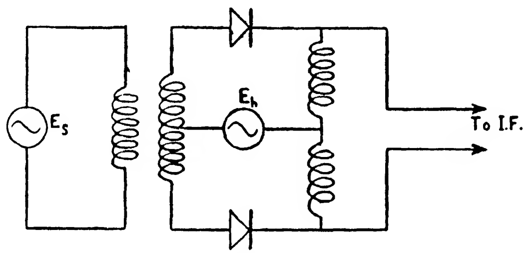
$$\eta = L_c(\eta_2 + \beta - 1) \qquad (4.105)$$

where L_c is the conversion loss, η_2 the noise factor of the I.F. amplifier, and β the noise temperature ratio of the crystal. Since L_c decreases with increasing crystal current and reaches an asymptotic minimum value, and β increases with increasing crystal current linearly, equation (4.105) shows that there will be a value for the crystal current where the noise factor is a minimum; if the crystal current is increased beyond this value, the noise factor will slowly deteriorate. The optimum noise factor is usually attained for crystal currents of about 0·5 mA, but the minimum is not critical, and any crystal current between 0·3 and 1·0 mA is usually satisfactory. The excitation power for a crystal current of 0·5 mA is about one milliwatt.

A receiver noise factor between 11 and 15 db can therefore be expected with a crystal mixer input circuit, assuming a 5 db noise factor for the I.F. amplifier. With suitable crystals, this performance can be maintained up to over 20,000 Mc/s.

4.4.4. *Local Oscillator Injection.* The main principle governing the injection of the local heterodyne signal into the mixer circuit is the requirement for very loose coupling. It is apparent from the schematic circuit of Figure 52A that a considerable amount of signal power will be sent down the local oscillator line, if this is tightly coupled to the circuit; and since the difference between the signal frequency and the local oscillator frequency is always relatively small at U.H.F., the signal power sent along the local oscillator line will not be reflected, but almost totally absorbed.

The demand for loose coupling from the local oscillator makes it necessary to supply much more power from the local oscillator than will actually be absorbed in the mixer. For crystal mixing, the local oscillator should be capable of delivering at least 20, preferably



E_s = signal source
E_h = local oscillator source

Fig. 53. Diagram of balanced crystal mixer circuit, used to obtain cancellation of local oscillator noise

50 milliwatts; and correspondingly more for diode and triode mixers.

A different consideration leads to the same demand for the highest possible power output from the local oscillator. If the noise sidebands in the local oscillator output are appreciable, they will intermodulate with the L.O. carrier in the mixer circuit, and the sidebands spaced away from the L.O. carrier by the intermediate frequency will appear as noise in the I.F. These noise sidebands from the local oscillator must therefore be attenuated as much as possible before reaching the mixer, and the simplest way of doing this is to increase the L.O. carrier power so that it can be substantially attenuated before reaching the mixer, while still giving an adequate output. In addition, the use of a high-Q resonant circuit in the oscillator, in conjunction with the use of a high intermediate frequency will attenuate the harmful noise components in the L.O. output before reaching the mixer.

If these precautions prove inadequate to overcome the local oscillator noise, a balanced mixer circuit is the final resort for the best possible sensitivity. This is shown schematically in Figure 53; it is seen that the I.F. noise components due to the local oscillator appear differentially in the output and should therefore cancel. Complete cancellation is difficult in practice, but with inherently noisy local oscillators such as reflex klystrons, the noise factor can be improved by several db. At the microwave frequencies, where this circuit is mainly used, the signal and oscillator frequencies are injected through a magic-T waveguide junction, which is electrically equivalent to the schematic shown.

4.4.5. *Local Oscillators.* The design of the local oscillator is governed by the same considerations as that of low-power transmitters, and in cross-band working the same circuit is actually used for both functions. Very little will therefore be said here.

Local oscillators can be classified from two different points of view; either as spot-frequency devices and wideband oscillators, or as free-running oscillators and frequency-controlled oscillators. Frequency-controlled oscillators can be further subdivided, according to whether direct low-frequency crystal control and frequency multiplication is used, or whether the oscillator is inherently free-running and has automatic frequency correction (A.F.C.) applied by a voltage derived from a frequency-discriminator circuit.

For wide-band free-running and A.F.C. controlled oscillators, triodes can be used up to 3000 Mc/s; but it is usual to change over to reflex klystrons or Heil tubes from about 2000 Mc/s upwards. Real wide-band operation over a 2 : 1 band is feasible at present up to 6000 Mc/s. Above this frequency, reflex klystrons are still used for A.F.C., but have a comparatively narrow tuning or control range. As already pointed out, the reflex klystron tends to be noisy owing to the interaction of the forward and returning electron streams, and the precautions outlined in Section 4.4.4 must be taken if receiver sensitivity is paramount.

A.F.C. with a triode needs a reactor valve; with the reflex klystron it is particularly simple in principle, as the A.F.C. bias can be directly applied to the reflector electrode. Extremely close frequency control can be achieved up to the highest frequencies.

In a wide-band receiver without a signal-frequency amplifier the mixer circuit can usually be ganged with the local oscillator circuit, since the selectivity of the mixer circuit is inherently so low that tracking errors can often be neglected. Where signal-frequency amplification is used with a receiver tuneable over a wide band, the local oscillator is best tuned separately, although some ingenious designs have been put forward for tracking a ganged receiver by intricate mechanical movements.

For spot-frequency and crystal controlled circuits, the low-power klystron is widely used, either as a direct oscillator, or as the final multiplier and amplifier of a signal derived from a low-frequency crystal oscillator by repeated frequency multiplication.

In free-running oscillators, whether wide-band or spot-frequency, the inherent frequency stability is an important factor; and an allowance has to be made in the I.F. band-width for the oscillator frequency tolerance.

Summing up, the main requirements for a local oscillator are adequate power output (50 mW), adequate frequency stability, low noise, adequate tuning range for the application in hand, and adaptability to A.F.C. where necessary.

REFERENCES AND BIBLIOGRAPHY

(1) H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," *Phys. Rev.*, 32, p. 110 (1928).
(2) J. S. Hey, T. W. Phillips and S. J. Parsons, *Nature*, 157, p. 296 (1946).
(3) W. Schottky, *Ann. d. Phys.*, 57, p. 541 (1918).
(4) W. Schottky and E. Spenke, *Wiss. Veroeff. Siemenswerke*, 16, p. 1 (1937).
(5) W. Schottky, *Ann. d. Phys.*, 32, p. 195 (1938).
(6) M. J. O. Strutt and A. van der Ziel, "Methods for Compensation of Shot Noise in Tubes and Circuits," *Physica*, 8, p. 1 (1941).
(7) "Microwave Receivers," *M.I.T. Radiation Lab. Series*, McGraw-Hill (1947).
(8) W. R. Ferris, "Input Impedance of Vacuum Tubes as U.H.F. Amplifiers," *Proc. I.R.E.*, 24, Jan. 1936.
(9) D. O. North, "Effect of Space-charge on Grid Impedance," *Proc. I.R.E.*, 24, Jan. 1936.
(10) D. O. North and W. R. Ferris, "Fluctuations Induced in Vacuum Tube Grids at High Frequencies," *Proc. I.R.E.*, 29, Feb. 1941.
(11) L. A. Moxon, "Noise Characteristics of Radar Receivers," *Journ. I.E.E.*, 43, Pt. IIIA, No. 6 (1946).
(12) M. J. O. Strutt, "Moderne Mehrgitterelektronenroehren," Springer, Berlin (1940).
(13) E. Hudec, "Kreuzmodulation und Eingangsrauschspannung," *E.N.T.*, 20, No. 5, May 1943.
(14) E. G. James and J. E. Houldin, "Diode Frequency Changers," *Wireless Engineer*, 20, Jan. 1943.
(15) B. Bleaney, J. W. Ryde and T. H. Kinman, "Crystal Valves," *Jour. I.E.E.*, 93, Pt. IIIA, No. 5 (1946).
(16) E. W. Herold and L. Malter, "Radio Reception at U.H.F.," *Proc. I.R.E.*, 31, Aug., Sept., Oct. 1943.

Refs. (11) and (16) are surveys of the whole field; the most detailed treatment will be found in ref. (7).

CHAPTER V

MODULATION TECHNIQUES

5.1. **General Considerations.** So many different modulation techniques are used nowadays in radio communication, that it may be well at the beginning of this discussion to set forth a few fundamental definitions. Let an un-modulated radio-frequency carrier be represented by the vector $A \sin(pt + \theta)$, the significance of which may be current, voltage or field strength. Then a general modulated carrier may be represented by $M(t)$, where

$$M(t) = F(t) \sin \{pt + f(t)\} \quad . \quad . \quad (5.1)$$

To the mathematician this would appear to be an unnecessarily complicated way of describing a modulated wave, but it is the most satisfactory way from the point of view of the engineer, since it introduces explicitly the unmodulated carrier frequency $p = 2\pi f_0$. In many cases $F(t)$ is a periodic function, and a simple example of this is where

$$M(t) = (1 + m_a \sin \mu_1 t) \sin \{pt + m_f \sin \mu_2 t\} \quad . \quad (5.2)$$

In this case the amplitude of the carrier is said to be modulated at a frequency $f_1 = 2\pi\mu_1$ and the phase of the carrier at a frequency $f_2 = 2\pi\mu_2$. Another mathematically simple example is given by

$$M(t) = a \sin(pt + \phi) \quad . \quad . \quad . \quad (5.3)$$

for values of t between t_1 and t_2, and

$$M(t) = 0 \quad . \quad . \quad . \quad . \quad . \quad (5.4)$$

for values of t between t_0 and t_1 and between t_2 and t_3, the phenomenon being periodic with a recurrence frequency $= \dfrac{1}{t_3 - t_1}$.

In this case the emission consists of a recurrence of square-topped pulses, each of length $t_2 - t_1$ and recurring at a frequency $1/(t_3 - t_1)$.

Neither phenomenon described by the above equations is exactly realizable physically, but sufficiently good approximations may be obtained to allow equations (5.2), (5.3) and (5.4) to be usefully employed.

For the purpose of this chapter it will be sufficient to define three broad classes of modulation.

(i) *Amplitude modulation*, where the amplitude of the carrier varies significantly and the phase (or frequency) does not vary significantly. This will be represented by

$$M_A = \{1 + F(t)\} \sin(pt + \phi) \quad . \quad . \quad (5.5)$$

(ii) *Phase or frequency modulation*, where the phase or frequency of the carrier varies significantly, but the amplitude remains constant. This will be represented by

$$M_F = \sin\{pt + f(t)\} \quad . \quad . \quad . \quad (5.6)$$

(iii) *Pulse modulation*, where the carrier is entirely absent for recurring periods, each of which contains many cycles of the frequency f_0. The amplitude of the carrier may vary from pulse to pulse, or the length of each pulse may vary, or the epoch at which each pulse occurs may vary. Any or all of these variations may be employed to impress the modulation on the carrier. It will be seen that (iii) contains phenomena associated with (i) and (ii), but physically, and particularly from the point of view of technique, pulse modulation is quite different from the other two classes.

A pure sinusoidal oscillation is a mathematical abstraction, and in communication engineering the investigator and designer are concerned with practical approximations. Some practical approximations to a sinusoidal oscillation are extremely good, but they are not perfect. Similarly, a pure amplitude modulated or pure frequency modulated wave is an abstraction. In practice any modulation technique produces a mixture of both. But by the proper design of equipment, both transmitting and receiving, the required characteristic of the modulated wave may be emphasized and utilized. Indeed, it may easily be shown that a wave of the type defined by equation (5.2) may be operated upon in a receiver so as to exhibit *either* the modulation frequency f_1 or the modulation frequency f_2 almost exclusively. It must therefore be noted that the intelligence (modulation) which is abstracted from a radio wave is as much a function of the receiver, as it is of the transmitter.

5.2. **Amplitude Modulation.** The general characteristics of amplitude modulation are too well known to require description, but one feature must be emphasized. The side-band pattern of an amplitude-modulated wave is completely determined by the frequencies present in the modulation spectrum. To take the simplest case, that of a sinusoidal modulation of frequency f_1 imposed upon a carrier of frequency f_0, the side-band pattern consists of two sinusoidal oscillations of frequencies $f_0 + f_1$ and $f_0 - f_1$, the amplitudes of which may be anything from zero to half the carrier, depending upon the depth of modulation. Amplitude modulation is unique in this respect, for in frequency (or phase) modulation and in pulse modulation, the side-band pattern is a function of the depth of modulation, and may spread between $f_0 - nf_1$ and $f_0 + nf_1$, where n is large, for a simple sinusoidal modulation of frequency f_1. As will be seen in the sequel, it is this curious limitation of amplitude modulation which makes it unsuitable for many communication purposes, and extremely unsuitable for use at U.H.F.

5.2.1. *Amplitude Modulation of a Common-grid Triode.* As at lower frequencies, two possibilities are open—low-level or high-level modulation—but these terms have rather different connotations when applied to common-grid triodes. Low-level modulation means modulating the cathode-bias voltage in a stage which absorbs an appreciable radio-frequency power and in which there is a significant negative feed-back. The modulated amplifier must, of course, be at the final radio-frequency, and the linear amplifier which in most cases would be required to raise the power level must itself be a common-grid stage. So far as the author is aware, such a low-level modulation technique has never been used outside the laboratory, but in certain cases it might have considerable value, particularly where the audio-frequency power drain is important.

High-level modulation of a common-grid amplifier stage is a well-known technique. Since the input and output powers are in series, it is necessary to modulate the two final stages. This may be done quite successfully, even where the penultimate stage is a frequency multiplier, the audio power being derived from two tappings on a single transformer. Usually, it is only necessary to modulate the driver stage about 30–50% and the output stage about 80%. Phase shift does not present any difficulty, and harmonic distortion can be kept below 2% without the addition of any complexity. The negative feed-back of the final stage is of assistance in this respect. Should the anode of the final stage be modulated alone, it is still possible (for poor quality services) to obtain 80–90% modulation with a harmonic distortion less than 10%. But in this case the power output of the transmitter does not increase in proportion to the depth of modulation. A stage which develops 10 watts unmodulated power, if it is 90% anode modulated without modulation of the driver, may develop 10 watts total power, corresponding to a carrier of about 7 watts. For some purposes this decrease in carrier power is acceptable, but it must be borne in mind that it is associated with a distortion level of the order of 10%.

Another important design requirement for a modulated common-grid stage is ample cathode emission to cope with the peaks of modulation. Present-day valves tend to be limited in their power output by cathode emission, and it is well to test the transmitter chain throughout at a carrier level equal to the required peak output, if 95% modulation without distortion is desired. This is particularly true of transmitters which are required to handle several speech channels simultaneously, as in multiplex working.

5.2.2. *Amplitude Modulation of a Klystron.* As in the case of the triode, modulation can theoretically be achieved by one of two means, roughly equivalent to high- and low-level operation. High-level modulation can be attempted by varying the accelerating potential V_0 by means of the modulating voltage. A modulating transformer may be interposed between the resonators and earth, the static accelerating voltage being applied negatively to the cathode. In Chapter III the output of a typical klystron was examined as a function of V_0, and the result is shown in Figure 37.

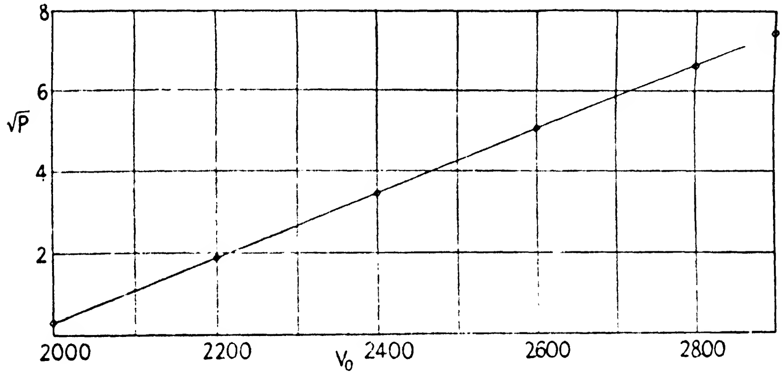


Fig. 54. Variation of the square root of the power output with the accelerating voltage in a klystron amplifier

Figure 54 corresponds to Figure 37, but exhibits $\sqrt{P}$ as a function of V_0, and shows that within the range 2000 to 2800 volts, the relation between the two is linear. If, therefore, V_0 is given the static value 2400 volts, and the audio-modulating voltage is 400 sin $\mu_1 t$, then the klystron is, so far as power is concerned, 100% modulated at the audio-frequency. It must be noted, however, that when V_0 is varied, the bunching parameter q is varied, which varies the phase difference between input and output. The input phase may be regarded as constant, in which case the variation of V_0 leads to phase modulation. Let the time in the drift-space, when V_0 has its mean value, be given by $T_0 = x/v_0$, as usual. Then the phase difference between input and output when the load is resistive is $\theta_0 = px/v_0$, where $v_0 = \sqrt{2V_0 e/m}$. Now when the voltage V_0 is modulated by the voltage $V_M \sin \mu_1 t$,

$$\theta = \frac{px}{\sqrt{\{V_0 + V_M \sin \mu_1 t\}}\sqrt{2e/m}} = \frac{k}{\sqrt{\{V_0 + V_M \sin \mu_1 t\}}}$$

$$= \frac{\theta_0}{\left\{1 + \frac{V_M}{V_0} \sin \mu_1 t\right\}^{\frac{1}{2}}}$$

Let $\frac{V_M}{V_0}$ be small, so that $\left(\frac{V_M}{V_0}\right)^2$, etc., can be neglected.

$$\text{Then} \quad \theta \fallingdotseq \theta_0 \left\{1 - \frac{V_M}{2V_0} \sin \mu_1 t\right\}$$

This represents a phase modulation of index $m = \frac{\theta_0 V_M}{2V_0}$ and the side-band spectrum is such that all frequencies differing from the carrier by more than

$$\delta f_0 = \frac{\mu_1}{2\pi}\left[1 + \frac{V_M \theta_0}{2V_0}\right]$$

can be neglected. Hence if the communication system is non-frequency selective over the band $f_0 - \delta f_0$ to $f_0 + \delta f_0$ the phase-modulation can be neglected, and an amplitude detector in the receiver will give a true modulation picture. For example, if $\theta_0 = 6\pi$, $V_M = 400$ and $V_0 = 2400$, then

$$\delta f_0 = \frac{\mu_1}{2\pi}\left[1 + \frac{\pi}{2}\right] = 2{\cdot}57\,\frac{\mu_1}{2\pi} = 2{\cdot}57 f_A$$

This is an exceedingly small phase-modulation at a micro-wave frequency, if f_A is in the audio region, and would not normally affect an amplitude-modulated system. High-level modulation of a klystron is, therefore, an attractive proposition. The modulator must supply power to the klystron, and, since the efficiency of the latter is low, the audio power consumption is relatively high; but there are many cases where this is not an important disability.

Low-level klystron modulation may be achieved by the use of an auxiliary electrode in the electron-gun structure. If the current from the cathode can be controlled by means of this auxiliary electrode, then it may be used to modulate the effective beam current I_0. As in the case of the control grid of a triode, however, such a method lends itself to distortion unless the depth of modulation is low. Theoretically, the method is free from the phase-modulation inherent in high-level modulation, so long as the variation of current does not seriously change the space-charge conditions in the beam. Also, this method does not require modulator power, since the control electrode is always negative with respect to the cathode, and draws no current. But, so far as the writer is aware, the practical realization of the necessary control has not so far been accomplished, and high-level modulation is preferred.

In general, klystron modulation suffers from the same disabilities as the modulation of the common-grid triode. In any given system, using a specified valve, the engineer can produce a satisfactory modulated amplifier: but the general principles of design, applicable to any valve cannot yet be formulated. Future development will depend upon the demand for this type of device. Should practice require engineering design, the latter will be arrived at by exact specification of the characteristics of the valve. But until more data are available regarding the relative performance of different individual systems, specification would not meet with the general approval required to give it effect.

Before leaving the subject of amplitude modulation, a word of warning should be added regarding the modulation of klystron self-oscillators. This is bad practice, as may be seen from the analysis of Chapter III. Variation of the beam current or the accelerating voltage or both produces pronounced phase-shift of the carrier, and this phase-shift, reflected back into the buncher resonator, produces serious distortion. In the extreme case of the reflex oscillator variation of the reflector voltage produces amplitude and frequency variation simultaneously, the frequency modulation giving rise to side-bands which may extend several megacycles on either side of the carrier.

5.3. **Phase or Frequency Modulation.** For the purpose of this general introduction phase and frequency modulation will not be distinguished, and both will be called by the more general term, phase modulation. The characteristics of this type of modulation are becoming as well-known as those of amplitude modulation, but to avoid misunderstandings the more important results will be recapitulated. The general modulated wave may be represented as in equation (5.6), or, taking $f(t) = m \sin \mu_2 t$ for simplicity, so that the modulation is sinusoidal, $M(t)$ may be written

$$M(t) = \sin pt \cos (m \sin \mu_2 t) + \cos pt \sin (m \sin \mu_2 t)$$

But

$$\cos (m \sin \mu_2 t) = J_0(m) - 2J_2(m) \cos 2\mu_2 t + 2J_4(m) \cos 4\mu_2 t - \ldots$$

and

$$\sin (m \sin \mu_2 t) = 2J_1(m) \cos \mu_2 t - 2J_3(m) \cos 3\mu_2 t + 2J_5(m) \cos 5\mu_2 t - \ldots$$

Whence

$$\begin{aligned} M(t) = {} & J_0(m) \sin pt \\ & + J_1(m) \{\cos (p + \mu_2)t + \cos (p - \mu_2)t\} \\ & - J_2(m) \{\sin (p + 2\mu_2)t + \sin (p - 2\mu_2)t\} \\ & - J_3(m) \{\cos (p + 3\mu_2)t + \cos (p - 3\mu_2)t\} \\ & + J_4(m) \{\sin (p + 4\mu_2)t + \sin (p - 4\mu_2)t\} \\ & + \ldots \\ & - \ldots \end{aligned} \qquad (5.7)$$

M(t), therefore, when analysed in terms of side-band pairs, contains an infinite number of the latter, the amplitudes of each of the r^{th} pair being given by $J_r(m)$.

The Bessel Function $J_r(m)$, considered as a function of r, m remaining constant, has its final maximum before asymptoting to zero at a value of r less than m. Figure 55 shows a typical case, where $J_r(20)$ is graphed as a function of r.

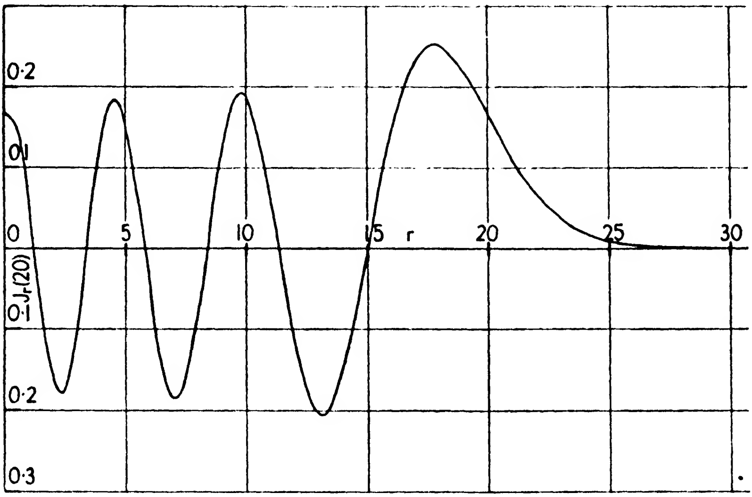


Fig. 55. Variation of $J_r(20)$ with r, showing that the function asymptotes to zero beyond $r = 20$

This means that $J_{m+2}^2(m) + J_{m+3}^2(m) + \ldots$ is always small compared with $J_0^2(m) + J_1^2(m) + \quad + J_m^2(m) + J_{m+1}^2(m)$ and so the power in the side-bands for which r is greater than $m + 1$ can be neglected. Hence, for all practical purposes, the F.M. spectrum can be taken to end at the pair of side-bands.

$$J_{m+1}(m)\{\sin(p + \overline{m+1}\,\mu_2)t + \sin(p - \overline{m+1}\,\mu_2)t\}$$

This is, in effect, the property which determines the minimum band-width necessary in the receiver to abstract all the intelligence from the transmission. For example, if $m = 3$, and the maximum value of $f_2 = \mu_2/2\pi$ is 3 kc/s, the receiver must have a minimum band-width of $\pm$ (3 + 1)3 kc/s = $\pm$ 12 kc/s, and the receiver discriminator must be linear over at least the same frequency interval. If m were changed to 20, the band-width would have to be increased from 12 to 63 kc/s for the same performance.

The total power in a frequency-modulated wave can be found by applying the usual laws to each side of equation (5.7). Let the total power be unity. Then

$$1 = J_0^2(m) + 2\{J_1^2(m) + J_2^2(m) + \ldots + J_r^2(m) + \ldots\} \quad (5.8)$$

since all cross-products obtained on squaring M(t) vanish on the average.

This is actually a physical derivation of a well-known deduction from the expansion in series of Bessel Functions. The significance to the theory of frequency modulation is that the side-band power

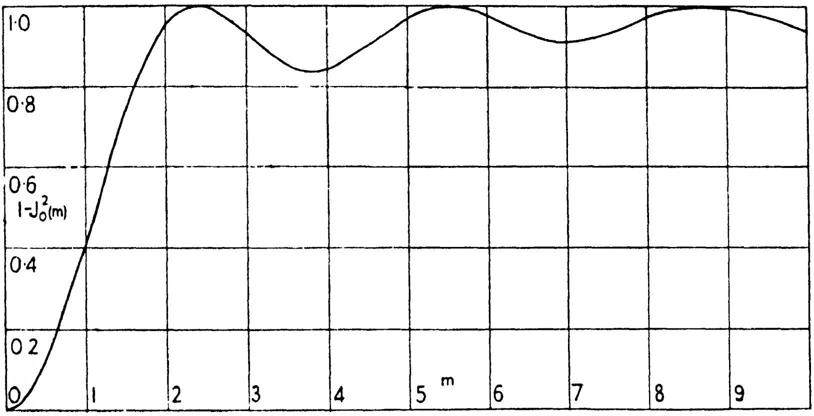


Fig. 56. Variation of the side-band power with the deviation ratio for a frequency modulated wave

is $2\{J_1^2(m) + J_2^2(m) + \ldots + J_r^2(m) + \ldots\}$ and the carrier power is $J_0^2(m)$.

Therefore, the ratio of the side-band power to the total power in the wave is $1 - J_0^2(m)$. This is shown as a function of m in Figure 56.

As was to be expected, the power in the side-bands increases very rapidly as the modulation index is increased from zero to unity. When an A.M. transmission is 100% modulated by a sinusoidal signal, the side-band power is 33% of the total. This occurs in F.M. when $m = 0{\cdot}85$. For all values of m greater than 0·85, the side-band power is greater, and indeed, at $m = 2{\cdot}4$, 5·52 and 8·65, $J_0(m) = 0$ and *all* the power is in the side-bands. Since the effective power of a transmitter to radiate intelligence is measured by the side-band power only, an F.M. transmitter is *a priori* more efficient for all values of m greater than 0·85.

It is now necessary to distinguish between phase and frequency modulation. First, if the carrier is sinusoidally phase-modulated, the modulated wave is $\cos \phi$ where $\phi = pt + a \cos \mu_2 t$. Here a is a constant for a given percentage modulation, and the phase-modulated wave is in more general terms

$$M_P(t) = R_P \cos (pt + a \cos \mu_2 t) \quad . \quad . \quad (5.9)$$

Second, if the carrier is sinusoidally modulated in frequency, the angular frequency is given by $q = p + b \cos \mu_2 t$ and the phase by

$$\phi = \int_0^t q dt = pt + \frac{b}{\mu_2} \sin \mu_2 t$$

Therefore, adding a phase-shift of $\pi/2$ for uniformity, the frequency-modulated wave is

$$M_F(t) = R_F \cos \left(pt + \frac{b}{\mu_2} \cos \mu_2 t \right) \quad . \quad . \quad (5.10)$$

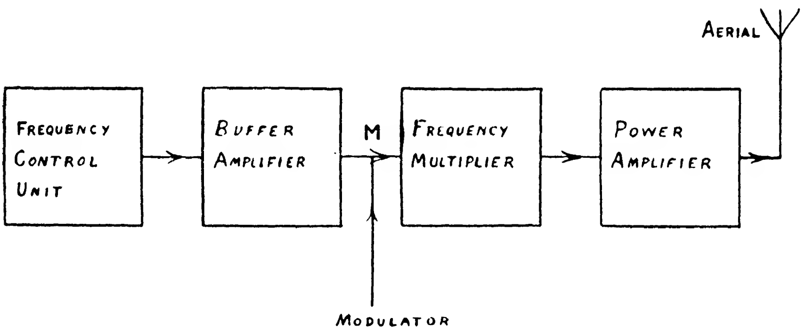


Fig. 57. Block diagram of a simple F.M. transmitter

and in this case $m = b/\mu_2$ is inversely proportional to the modulating frequency.

The techniques for the production and detection of phase- and frequency-modulated waves are very similar, and the similarity is complicated by modern methods of pre-emphasis and de-emphasis. It must always be remembered, however, that the two waves behave differently during propagation, particularly if there are two or more possible paths.

5.3.1. *Phase or Frequency Modulation of a U.H.F. Crystal Controlled Transmitter.* Ultra high frequency transmitters may be designed on orthodox principles and according to the block diagram of Figure 57, in which the modulator is omitted, since it may take one of several forms.

Without using velocity modulation devices, transmitters of this type may be built for frequencies up to 3000 Mc/s, so long as the power output required is not more than a fraction of a watt. By using klystron multipliers and amplifiers, the principle may be extended to frequencies of the order of 10,000 Mc/s. In the next chapter frequency control techniques will be discussed in detail, but for the purpose of this investigation it will in the first instance be assumed that the frequency control unit is a crystal-controlled oscillator. Then, if the frequency of the latter is chosen so that the frequency multiplication in the transmitter chain is adequate, it is possible to phase-modulate the carrier at M by means of a balanced modulator or one of its variations. This is the method originally introduced by Armstrong,[1] and it gives very stable operation with complete control of all variables. Where it is practicable, the Armstrong system is very satisfactory.

A more detailed block diagram of an Armstrong modulator is shown in Figure 58, where the crystal oscillator, followed by its buffer amplifier, is used to feed the carrier to the balanced modulator. The side-band output from the latter is given a 90° phase-shift relative to the carrier before being mixed with the latter in a linear device. The output from this last stage is a phase- and amplitude-modulated wave which must then have its frequency multiplied many times in order to transfer the power to the modulation side-bands. The small percentage of amplitude modulation (which is not increased by frequency multiplication) may either be neglected altogether or may be obliterated by means of a limiter in the chain.

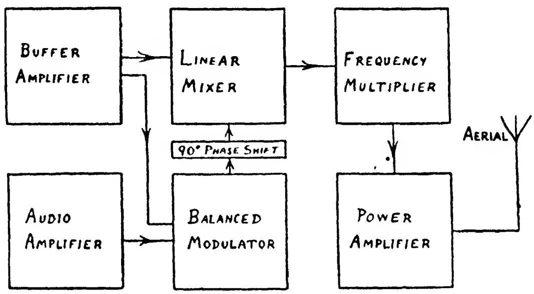


Fig. 58. Block diagram of a transmitter employing an Armstrong modulator

It is interesting to calculate how much multiplication of the frequency of the modulated carrier is necessary to achieve a certain phase or frequency deviation without distortion, and it is a valuable result, since Armstrong did not envisage the use of this modulation technique for narrow-band working, and therefore tends to give the impression that a very high multiplication factor is always necessary.

Let the carrier at the crystal oscillator frequency be represented by $\sin pt$. Then the modulated carrier is $(1 + m' \cos \mu t) \sin pt$ and the output from the balanced modulator is the suppressed carrier modulation $m \cos \mu t \sin pt$. If the phase of the carrier is changed by 90° and it is added to the side-bands in a linear device, the resulting wave is

$$M(t) = \cos pt + m \cos \mu t \sin pt$$

Multiply and divide throughout by $\sqrt{1 + m^2 \cos^2 \mu t}$, then

$$\mathrm{M}(t) = \sqrt{1 + m^2 \cos^2 \mu t}\left\{\frac{1}{\sqrt{1 + m^2 \cos^2 \mu t}} \cos pt + \frac{m \cos \mu t}{\sqrt{1 + m^2 \cos^2 \mu t}} \sin pt\right\}$$

$$= \sqrt{1 + m^2 \cos^2 \mu t} \cos \{pt - \tan^{-1} (m \cos \mu t)\}$$

Now let m be small, so that

$$\tan (m \cos \mu t) \risingdotseq m \cos \mu t = \theta \quad . \quad . \quad . \quad (5.11)$$

Then $$\mathrm{M}(t) \risingdotseq \cos (pt - m \cos \mu t)$$

which is a phase-modulated wave of index m. The amplitude modulation is given by $(1 + \frac{1}{2}m^2 \cos^2 \mu t) \cos pt$ and is very small if m is small.

If what is required is a frequency-modulated wave, the input from the audio stage to the balanced modulator is taken from a capacitance in series with a resistance according to the usual procedure: in this case $m = k/\mu$. So long as θ is small in equation (5.11), a rough guide to the distortion to be expected is that a difference of $n\%$ between θ and $\tan \theta$ leads to a harmonic distortion of $n\%$. For the purpose of this calculation $\tan \theta$ can be taken equal to m, i.e. equal to k/μ.

Then for 1% distortion, k/μ must be less than $\frac{1}{8}$ and for 10% distortion, k/μ must be less than $\frac{1}{2}$. Suppose now that the highest audio-frequency is μ_{H} and the lowest audio-frequency is μ_{L}. Then the maximum value of m is k/μ_{L}. This must be less than a certain fraction $1/s$. Therefore k must be less than μ_{L}/s at the carrier frequency. If a deviation ratio of r is required from the transmitter, then $k/\mu_{\mathrm{H}} = r$ at the final frequency, or $k_{\mathrm{F}} = r\mu_{\mathrm{H}}$. Therefore the multiplication factor k_{F}/k must be greater than or equal to $sr\mu_{\mathrm{H}}/\mu_{\mathrm{L}}$.

To exemplify, let the distortion level be specified as less than 10%, and the deviation ratio as 2 for a speech-band covering the audio range 300–3000 c/s. Then $s = 2$, $r = 2$, $\mu_{\mathrm{H}}/\mu_{\mathrm{L}} = 10$, and the multiplication factor must be greater than $2 . 2 . 10 = 40$. On the other hand, if the distortion level is to be less than 1% and the deviation ratio is again 2, but the audio-frequencies ranging from 50–15,000 c/s, the multiplication factor must be at least 3600. This example corresponds to a broadcasting system, employing a deviation of 30 kc/s.

It should be noted that the above analysis is concerned only with the multiplication factor required. It takes no account of other forms of distortion which may be serious. In particular, care must be taken with the linear mixer when the Armstrong modulator is used.

If the frequency multiplication available is not sufficient to allow a balanced modulator to be used, one other modulation technique is open. A variable frequency oscillator must be incorporated in the system, its *mean* frequency being maintained at a constant value by means of automatic frequency control from a crystal oscillator and an associated discriminator. This variable frequency oscillator can then be modulated in the usual manner, and since reactance modulation allows relatively large frequency or phase excursions, the overall frequency multiplication need not be so large. This method is also the obvious one to use if, for any reason, a variable frequency oscillator is already incorporated in the transmitter chain. As will be seen in the next chapter, this is a common situation where a complex F.C.U.* is employed to give a multiplicity of transmitting frequencies.

5.3.2. *Phase or Frequency Modulation of a Reflex Klystron Oscillator.* As will be seen in the next chapter a high stability U.H.F. transmitter or receiver can be made in which the self-oscillator is at the radiated frequency, even when the latter is as high as 10,000 Mc/s. This is done by making use of the dependence of the frequency of the oscillator upon operating voltages, and advantage can be taken of the same characteristic to impress upon the emitted wave a frequency or phase modulation.

In Chapter III, paragraph 3.2.1, it was shown that the equation governing the frequency variations of a reflex klystron is (3.13)

$$\delta f = \frac{f_0}{2Q_E}\tan\phi \qquad . \quad . \quad . \ (5.12)$$

where (3.11) $$\frac{2dv_0 p}{(V_0 + V_R)\,e/m} = 2\pi\left(n - \tfrac{1}{4} + \frac{\phi}{2\pi}\right) \qquad . \quad . \ (5.13)$$

If V_0 is maintained at a constant value and V_R is varied by a small amount, equation (5.13) may be written

$$\frac{2\pi f A}{(V_0 + V_R + \delta V_R)} = 2\pi\,(n - \tfrac{1}{4}) + \phi \qquad . \quad . \ (5.14)$$

where $A = \dfrac{2dv_0}{e/m}$ is a constant for the valve, and

$$\frac{2\pi f_0 A}{V_0 + V_R} = 2\pi\,(n - \tfrac{1}{4}) \qquad . \quad . \quad . \ (5.15)$$

Eliminating $V_0 + V_R$ between equations (5.12), (5.14) and (5.15) and remembering that δV_R and therefore δf are small,

$$\delta f = -\frac{\pi\,(n - \tfrac{1}{4})^2}{AQ_E}\,\delta V_R \qquad . \quad . \quad . \ (5.16)$$

* Frequency Control Unit.

This equation, in which the parameter n (an integer) replaces the parameter V_R, exhibits the manner in which the gradient df/dV_R varies with the mode of oscillation and with the Q_E of the system.

Now let $\delta V_R = a \cos \mu t$. Then $f = f_0 + \delta f$

and
$$f = f_0 - \frac{\pi (n - \frac{1}{4})^2}{AQ_E} a \cos \mu t$$

whence the modulated wave is

$$M(t) = B \sin \left(p_0 t - \frac{2\pi^2 (n - \frac{1}{4})^2}{\mu AQ_E} a \cos \mu t \right) \quad . \quad (5.17)$$

Inserting reasonable values for the parameters in equation (5.17), for example, $n = 3$, $a = 10^8$ e.m.u., $\mu = 20{,}000$ r/s, $d = 0{\cdot}5$ cm., $Q_E = 400$, $V_0 = 3{\cdot}10^{10}$ e.m.u., and $e/m = 1{\cdot}8 . 10^7$ e.m.u., the deviation ratio comes out at about 30, corresponding to a frequency deviation of about 100 kc/s. This is achieved by means of a reflector voltage variation of amplitude 1 volt, which, in general, would produce a negligible effect upon the amplitude of the U.H.F. wave. A very pure frequency modulation therefore results. If, however, a were increased to 10^9 e.m.u., corresponding to 10 volts, a significant amplitude modulation would be associated with the frequency modulation, and steps would have to be taken to eliminate this in the receiver.

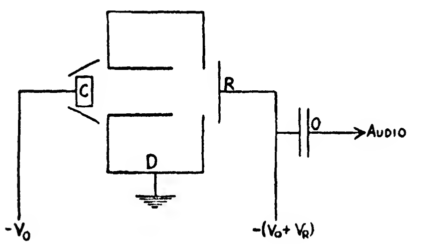


Fig. 59. Illustrating the method of applying frequency modulation to a reflex oscillator

The circuit employed to produce frequency modulation of a reflex oscillator is very simple, and is shown in Figure 59, where C is the cathode, D is the U.H.F. resonator, and R is the reflector. The steady potentials are applied as shown, but the audio voltage is impressed *via* the condenser O. It should be noted that little or no audio power is required. When automatic frequency control is simultaneously applied to such a self-oscillator, relatively narrow-band modulation can be imposed with a surprisingly high signal-to-hum ratio. The author has used such a transmitter at 3000 Mc/s with a deviation of 30 kc/s, the upper audio-frequency being 15 kc/s.

5.3.3. *Klystron Amplification of a Phase- or Frequency-modulated Wave.* A complete transmitter, the oscillator of which is a reflex klystron, requires a buffer or power amplifier between the latter and the aerial to prevent frequency variations due to variable load impedance. Care must be taken, however, if the amplifier is a klystron, for, as has been shown in paragraph 5.2.2, variations in the supply voltage produce in this velocity modulation device variations in the phase between input and output similar in quality to the modulation variations in the reflex oscillator. Hence the klystron amplifier power supply should be just as stable as the oscillator power supply, and since the voltages and currents are usually much larger than in the case of the oscillator, this may present considerable difficulty.

The unwanted effect, due to voltage variations in either, becomes less important as the modulation deviation increases, but it makes the use of an amplifier difficult where narrow-band frequency modulation is employed. For example, it is shown in paragraph 5.2.2 that the phase-modulation index is given by

$$m = \frac{\theta_0 V_M}{2V_0},$$

where θ_0 is the mean phase difference between the input voltage and the output current, and V_M is the amplitude of the voltage fluctuation in V_0. Let $\theta_0 = 6\pi$, $V_M = V_0/10$, so that there is a 10% ripple in the supply voltage. Then $m = 6\pi/20 \fallingdotseq 1$. If the hum level is to be 40 db below the modulation, the minimum modulus of the frequency deviation, D/f_A, where f_A is the maximum audio-frequency, must be 100. That is, D must be $100 f_A$. For a speech circuit f_A may be taken to be 3 kc/s. Therefore, to have the hum level 40 db below the 100% modulation level, the deviation must be 300 kc/s.

The author has experienced this difficulty in attempting to use a klystron power amplifier at 3000 Mc/s as the final unit in a crystal-controlled transmitter chain. The klystron was driven by means of a triode frequency multiplier, transforming the carrier at 1000 Mc/s to a carrier at 3000 Mc/s. Frequency modulation of deviation 6 kc/s was applied by means of the Armstrong method at a carrier frequency of 4·125 Mc/s. The output from the final multiplier gave a 100% modulation/hum ratio exceeding 50 db. The addition of a klystron amplifier with a simple power supply reduced this ratio to 20 db. It was very difficult to bring the hum level down to its previous value.

5.4. **Pulse Modulation.** Of the three types of modulation introduced in paragraph 5.1 pulse modulation is the most modern and the most complex. Speaking metaphorically, amplitude modulation has but one degree of freedom. A carrier, 100% modulated by a sinusoidal signal, is a definite physical phenomenon, both as regards its side-band spectrum and its side-band power. Using the same metaphor, phase or frequency modulation has two degrees of freedom. The carrier may be regarded as 100% modulated by a sinusoidal signal for an infinite number of deviation ratios, although, mathematically, only one parameter appears. Pulse modulation has at least three degrees of freedom in any one of its various forms. A 100% modulated wave has a double infinity of variations of pulse length and pulse repetition frequency.

Without taking into account coding systems, such as morse, there are three basic forms of pulse modulation which may be called

(i) amplitude modulation (P.A.M.);
(ii) length modulation (P.L.M.);
(iii) position modulation (P.P.M.).

As is suggested by the names, in (i) the modulation is used to vary the amplitude of the pulse, in (ii) to vary the pulse length, and in (iii) to vary the pulse position. In each case it is assumed that the unmodulated, pulsed carrier consists of a recurrence of pulses of equal shape and height at equal time intervals.

Of the three variants, amplitude modulation is the simplest to understand and was the first to be used, but more recent practice prefers length or position modulation. Pulse amplitude modulation gives no significant improvement over C.W. amplitude modulation so far as signal/noise ratio is concerned. It is more convenient than the latter in certain cases, particularly in the micro-wave region, where C.W. valves are not so powerful as the pulsed variety. No more will be said of P.A.M. and attention will be focused on pulse systems which are inherently better in some respects than the older and more conventional types.

5.4.1. *Pulse Length Modulation.* The modulation process may be followed from the idealized Figure 60.

Imagine that the transmitter in an unmodulated condition produces recurring pulses of the type shown in line (*a*), the length of each pulse and the repetition frequency being constant. Let an audio sinusoidal modulation, represented by the curve of line (*b*) be applied to the transmitter. Let the circuits be so arranged that the length of the pulse is made equal to $a + b \sin \mu t$, where the modulation is $\sin \mu t$ and a and b are constants. Then the transmitter, when modulated, emits recurring pulses of the length and position shown in line (*c*). If the receiver contains integrating demodulator circuits, such that the output voltage is proportional to the area of each pulse, then the former will reproduce the modulation according to the dotted envelope of line (*d*).

This is the simple, qualitative theory of P.L.M. and it is now necessary to investigate how far such a system can be used to transmit speech or other intelligence, and how the modulation can be achieved in a U.H.F. transmitter. The first question which arises is how often per cycle the modulation must be sampled to give a certain degree of intelligibility. For the purpose of this argument it will be assumed that a single speech channel is desired, and it may be noted that the results apply not only to P.L.M. but, as a first approximation, to P.P.M. as well. As might be expected, it is found that the pulse repetition frequency (P.R.F.) must be at least

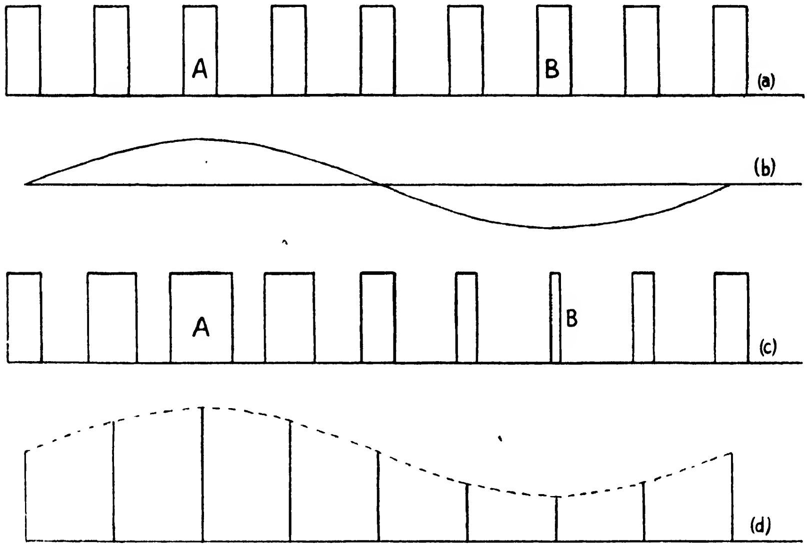


Fig. 60. Diagram illustrating the principle of pulse length modulation

twice the highest audio-frequency for good intelligibility. Surprisingly, if the P.R.F. is increased beyond this point, the gain in intelligibility is slight in the absence of interference. This is not at all obvious from a study of Figure 60, where, if the P.R.F. is twice the audio-frequency, the picture would be reduced under the most favourable conditions to the pulses marked A and B only. The result becomes a little more understandable if the demodulating process is considered. The pulses A and B, recurring, would produce a signal after the demodulator, consisting of approximately triangular humps and dips, as shown in Figure 61. To avoid the effect in the figure of the P.R.F., the latter has been taken to be rather greater than twice the audio-frequency.

The wave-form of Figure 61 contains the audio signal as the fundamental in its Fourier components, and so, if it is passed through an appropriate filter, excluding the P.R.F. and its harmonics, but accepting audio-frequencies, the modulation will reappear.

It should be noted, however, that although the audio filter after the receiver demodulator rejects all frequencies higher than a certain value, the receiver up to and including the demodulator must accept as much of the pulse side-band spectrum as is practicable, for the modulation is superimposed thereon. It is, therefore, worth while to examine the spectrum of the unmodulated recurrence of pulses more closely. Let the P.R.F. be f_P and the pulse length be p sec., so

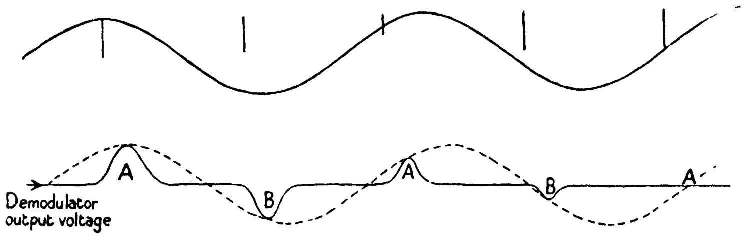


Fig. 61. Illustrating the effect of a pulse repetition frequency slightly greater than twice the modulating frequency

that the duty cycle of the pulse modulator is $pf_P = r$. Then the recurrence of square-topped pulses of unit amplitude may be represented by the Fourier series

$$P(t) = pf_P + \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{\sin(n\pi pf_P)}{n} \cos(2\pi nf_P t)$$

Rewrite this in terms of the duty cycle, r,

Then $$P(t) = r\left[1 + 2\sum_{1}^{\infty} \frac{\sin(n\pi r)}{n\pi r} \cos(2\pi nf_P t)\right]$$

The modulated carrier of frequency f_0 can therefore be written

$$P(t)\cos(2\pi f_0 t)$$

$$= r\left[\cos(2\pi f_0 t) + 2\sum_{1}^{\infty} \frac{\sin(n\pi r)}{n\pi r} \cos(2\pi nf_P t)\cos(2\pi f_0 t)\right]$$

$$= r\left[\cos 2\pi f_0 t + \sum_{1}^{\infty} \frac{\sin(n\pi r)}{n\pi r} \{\cos 2\pi(f_0 + nf_P)t + \cos 2\pi(f_0 - nf_P)t\}\right] \quad (5.18)$$

Taking the mean square value of each side as a measure of the power, and multiplying by $2/r$,

$$1 = r + 2r \sum_{1}^{\infty} \frac{\sin^2 (n\pi r)}{n^2\pi^2 r^2} \quad . \quad . \quad (5.19)$$

a result which may be derived in pure mathematics from the theory of Infinite Products.

If, therefore, the total power in the pulsed carrier is taken as unity,

the side-band power is $2r \sum_{1}^{\infty} \frac{\sin^2 (n\pi r)}{n^2\pi^2 r^2}$

or
$$\frac{\text{Side-band Power}}{\text{Total Power}} = 1 - r \quad . \quad . \quad (5.20)$$

This result may be compared with that for a frequency-modulated wave, derived in paragraph 5.3, but it must be noted that whereas the F.M. result referred to the modulation side-bands, the equation (5.20) above refers to the side-bands due to the unmodulated pulses.

The receiver must have a wide enough pass-band to accept a reasonable fraction of the side-band power, since each is modulated at the audio-frequency. If the "reasonable fraction" is taken as $a\%$, then mathematically

$$r + 2r \sum_{1}^{m} \frac{\sin^2(n\pi r)}{n^2\pi^2 r^2} \text{ must be} > a/100$$

$$\text{or} \quad \sum_{1}^{m} \frac{\sin^2(n\pi r)}{n^2\pi^2 r^2} > \frac{a}{200\,r} - \tfrac{1}{2}$$

an equation which defines m, and therefore the band-width of the receiver, $2mf_P$.

The function

$$\frac{\sin^2 (n\pi r)}{n^2\pi^2 r^2} \text{ may be written } \frac{\sin^2 x}{x^2}$$

and this falls slowly from unit value at $x = 0$ to zero at $x = \pi$. Clearly, we are not interested in values of x greater than $\pi/2$, from which the maximum value of the product nr may be taken as $1/2$, or $m_{max} = 1/2r$ and $2mf_P = f_P/r = 1/p$. The band-width of the receiver should therefore be at least the inverse of the pulse-length employed.

To investigate the spectrum of the modulated pulses is difficult. It cannot be assumed, as was done in early investigations, that the spectrum is obtainable by writing one of the parameters of equation (5.18) in the modulated form. For example, it is not satisfactory simply to write $r = r_C(1 + m \cos \mu t)$ in order to represent the audio modulation in P.L.M., nor is it correct to write

$$f_P = f_{PC}(1 + m' \cos \mu t)$$

for P.P.M. A more fundamental examination is required. However, so far as the engineer is concerned, it may be assumed that the modulation is carried on all the side-bands, and so there may be a double infinity of difference frequencies present in the spectrum.

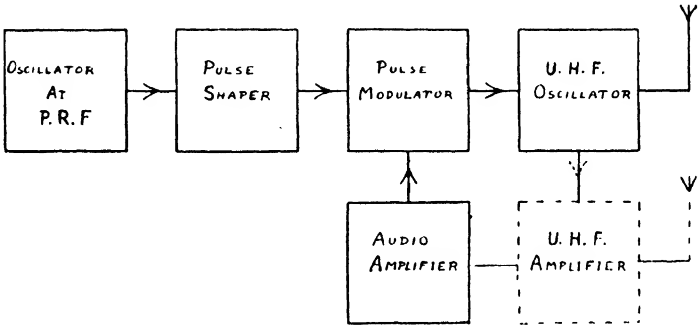


Fig. 62. Block diagram of a P.L.M. transmitter

5.4.1.1. *Circuit Techniques for P.L.M.* Circuit techniques for pulse modulation are discussed with an adequate bibliography by Cooke, Jelonek, Oxford and Fitch,(2) and this may be supplemented by an examination of standard radar pulse circuits which have many features in common. The block diagram of a P.L.M. transmitter is shown in Figure 62, where the radio-frequency units are left rather indeterminate.

The oscillator which determines the P.R.F. may be a crystal-controlled multi-vibrator or any similar unit. This may be followed by a pulse-shaping circuit which preferably gives a saw-tooth pulse with a rapid recovery. This is applied to the modulator simultaneously with the amplified audio output. The latter determines the output pulse length by defining the conducting period of the modulator. This valve is arranged to operate into anode saturation, so that the amplitudes of the pulses are equal and only their lengths vary.

The saw-tooth pulse shape helps to make the modulation linear, although other considerations, such as weight and size and freedom from jitter may favour the use of a sine-wave. The U.H.F. part of the transmitter is generally very simple. In most existing transmitters, it is a self-oscillator in which the pulse is a simple on-off switch. This is one of the important advantages of pulse modulation —that it does not require great carrier frequency stability. For that reason the modulation of a self-oscillator is feasible. On the other hand the modulation might be applied as shown by the dotted lines of Figure 46 to a power amplifier stage. An important difficulty here is the fact that common-grid triode amplifier stages cannot easily be switched off by biasing back the grid voltage. Power is absorbed in the cathode-grid circuit, and this may be prohibitive where the pulse-length is small.

It is perhaps worth while to recall at this point a feature of pulse modulation which is fundamental. The cathode of a thermionic valve is capable of a much greater emission per unit area in the pulsed condition than it can provide C.W. A reasonable figure for C.W. operation is 100–200 mA/sq. cm., while the same valve, when operating on a pulsed duty cycle of the order of 1/10 will readily give 1–2 A/sq. cm. Since cathode emission may be the limiting factor in a U.H.F. transmitter, this is a most important consideration.

5.4.2. *Pulse Position Modulation.* This term embraces several modulation techniques which are variously described as pulse frequency- and pulse phase-modulation. The basic feature is, however, stressed by the word "position." The amplitude and width of each pulse is a constant; the variable is the position of the pulse on the time scale.

Much of what has been said regarding P.L.M. is equally applicable to P.P.M. For example, it is equally true that the mean P.R.F. must be about twice the highest modulation frequency for small distortion; that the band-width of the receiver should be at least $1/p$, where p is the pulse length, for adequate and distortion-free demodulation; and that this type of modulation is superior in overall performance to P.A.M. A diagram similar to Figure 60 may be drawn for P.P.M., showing that, whereas in the case of P.L.M. the integrated effect of the pulse is the important factor, in the case of P.P.M. only the position of the leading edge of the pulse is significant. *A priori* a system depending upon the positioning of an abrupt switch-on of the carrier possesses some of the characteristics of frequency modulation; in particular, the demodulator, as in F.M., must be some form of discriminator. In practice it includes its own local comparison pulse.

5.4.2.1. *Circuit Techniques for P.P.M.* A block diagram of a simple P.P.M. transmitter is shown in Figure 63. The speech is employed to vary the frequency of the oscillator by means of a reactor valve. The frequency modulated wave is then changed to the desired length and amplitude by means of shaping circuits, and the resultant pulses are applied as before to a U.H.F. oscillator or power amplifier. This would be a pulse frequency-modulated transmitter, since the modulation of the P.R.F. is given by

$$f_P(1 + m_f \cos \mu t),$$

where $m_f \cos \mu t$ is proportional to the modulation amplitude.

In general, pulse phase-modulation requires the same circuits as P.L.M. to produce the variable position of the leading edge of the pulse, but additional circuits must be interposed between the modulator and the oscillator (Figure 62) in order to differentiate

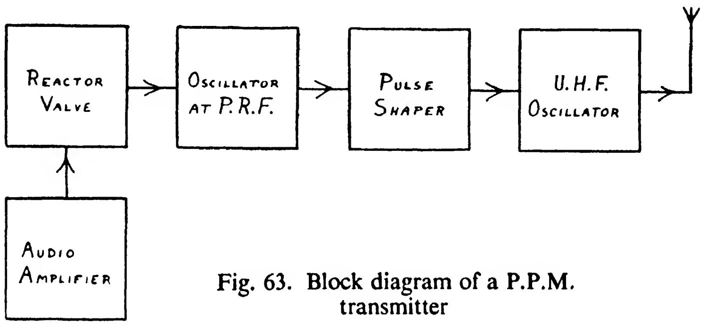


Fig. 63. Block diagram of a P.P.M. transmitter

and then shape the pulses to uniform size at non-uniform spacing. In this case if ϕ is the mean pulse phase in terms of the mean P.R.F., so that $\phi = 2\pi f_P t + \phi_0$, then the modulation changes ϕ to

$$2\pi f_P t + m_P \cos \mu t + \phi_0.$$

The demodulating circuits of a P.P.M. system do not suggest themselves as readily as in the case of P.L.M. Several methods have been employed, two of which are of generic interest. First, there is the simple comparison method, whereby the position modulated pulse is applied to one grid of a pentode, a pulse of constant P.R.F., locally generated, being applied to the other. This combination of voltages is made to produce a saturated anode current for a time interval equal to the difference in time between the leading edges of the two sets of pulses. This converts P.P.M. to P.L.M. An integrating circuit in conjunction with filters then restores the audio modulation. Alternatively, the two pulse trains, one of which is again generated locally, can be made to give an output pulse the amplitude of which is proportional to the interval between the leading edges. This output pulse is amplitude modulated, and may be further demodulated in the orthodox manner.

5.4.3. *Pulse Multiplex.* It is unusual to employ pulse modulation for single channel working, where it offers few advantages over the older methods, but it is very well suited for multiplex. If a pulse-length of the order of 1 microsecond is employed, 500,000 pulses can be sent per second. If each channel is associated with a P.R.F. of 10 kc/s, about 50 channels can be accommodated with an associated receiver band-width of a few megacycles per second. Such a system would be difficult to handle, and in practice twenty-four channels is a more common number. This requires twenty-five

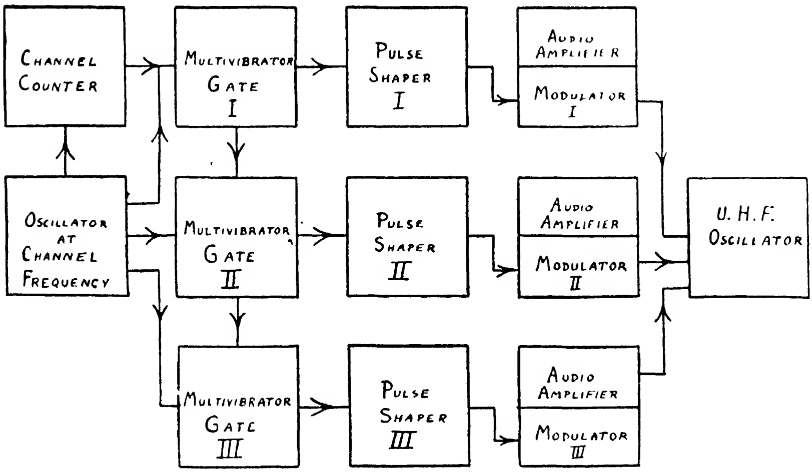


Fig. 64. Block diagram of a modulator for multiplex pulse communication

pulses, each of which recurs at a frequency of 10 kc/s. Mathematically, such a pulse-modulated multiplex system is very similar to the more orthodox frequency division system. The supersonic sub-carrier employed in the latter is in some ways analogous to the sub-carrier frequency f_1, in the former. In short, the degrees of freedom are the same in both systems. Pulse modulation multiplex equipment employs simpler and cheaper components than sub-carrier multiplex equipment, the costly precision filters of the latter being no longer necessary. Typical circuits will be discussed in the next paragraph. On the other hand, sub-carrier multiplex is essentially superior to pulse multiplex (sometimes called "time division multiplex"), provided that the proper precautions are taken against cross-modulation. The superiority lies in the theoretically achievable signal/noise ratio for a given power and band-width. In particular, an F.M. sub-carrier multiplex system is usually about 20 db better than a pulse multiplex system for equal numbers of channels. In a later chapter the relative advantages and disadvantages of these different systems will be more closely studied.

5.4.3.1. *Circuit Techniques for Multiplex Operation.* The simplest form of modulator is probably that which modulates each channel separately and then combines all the pulses to switch the carrier on and off. The block diagram of Figure 64 illustrates one form of this, only the essential features being presented.

In explanation, suppose that there are n channels. Then the channel oscillator at the left-hand side operates at a P.R.F. of $(n + 1)f_A$, where f_A is the frequency with which a given channel recurs. One output from this oscillator is fed to an adding circuit which "triggers" after $(n + 1)$ pulses. This is the "channel counter." The other output from the oscillator is fed in parallel to the chain of n multivibrators in such a manner as to stop the latter from generating pulses. When the counter operates on the synchronizing pulse, it triggers the first multivibrator and the pulses propagate along the chain. These pulses from the multivibrators can then be used to produce the separate channel pulses which are to be modulated. Finally, all the pulses are applied in turn to the U.H.F. oscillator.

Another method of generating pulses at time-spaced intervals in a definite cycle is to make use of a delay network. Imagine a delay line consisting of n units. If a multivibrator initiates a pulse at one end of the line, a distorted and delayed version of this may be fed from the line to n separate pulse-forming units, and the pulse eventually emerging from the delay line may be used to re-trigger the multivibrator. This is a rather more elegant "gating" system than the first and it has more promise of development.

A more complex form of modulator is one in which the modulation for each channel is applied to a common pulse unit. This demands a special equipment in which the pulses are generated, separated in space as well as in time, but in one unit. A modified type of cathode-ray tube, sometimes known as the "cyclophon," has been developed to do this, and the method is used in U.S.A. Demodulating circuits in the receiver reverse the actions described above for the transmitter. The gating pulses, by varying control electrode potentials, determine through which demodulating circuits the incoming pulses will pass. Thereafter, the latter are demodulated in the manner previously described.

REFERENCES

(1) *Proc. I.R.E.*, 24, May, 1936.
(2) *J.I.E.E.*, Pt. IIIA, 11 p. 83 (1947).

CHAPTER VI

FREQUENCY CONTROL

6.1. **General Considerations.** Having studied the various modulation techniques available at U.H.F., it is now necessary to investigate the quality of frequency control desirable in each case. The permitted tolerance in the frequency stability of a transmitter or receiver is a function of many variables, and from the outset it will be assumed that the discussion is concerned only with technical considerations. It may very well be that the frequency tolerance of a transmitted carrier is defined by purely political decisions. Such matters of national or international agreement will be neglected.

Consider two entirely different modulation techniques, such as amplitude modulation on the one hand and pulse modulation on the other. To emphasize the difference between them let both be used for single-channel speech communication, but let the pulse-modulation system employ a pulse length of 1 microsecond. To add an air of verisimilitude let the carrier frequency in both cases be 10,000 Mc/s. Then the modulation band-width in the A.M. case is $\pm$ 3 kc/s, while in the P.M. case it is $\pm$ 1 Mc/s. If the frequency control applied to the A.M. system were such that there were no variations in frequency greater than 1 part in 10^8, the receiver band-width need only be $\pm$ 3 kc/s. If the frequency control were relaxed to accommodate variations as large as 1 part in 10^4, the receiver band-width would require to be $\pm$ 1 Mc/s. This second A.M. receiver would be much less sensitive than the first; indeed, the first would detect a signal 25 db below the minimum detectable by the second.

Now consider the same conditions of frequency control applied to the P.M. system. The band-width in both cases would be of the same order of magnitude, being $\pm$ 1 Mc/s with the accurate control and $\pm$ 2 Mc/s with the relaxed control. The sensitivity difference between these two would be less than 1 db.

Here there are immediately presented two important aspects of frequency control, (*a*) that improved control may mean improved sensitivity, and (*b*) that there is a definite limit to the operation of this rule, the limit depending upon the nature of the modulation. From this emerges the general technical principle governing the design of control mechanisms for all radio links in which the ultimate sensitivity of the system is an important factor. This principle may be stated as follows: So far as ultimate sensitivity is concerned, the frequency control should be such that the band-width in the receiver required to contain the modulation side-bands is not very much less than the total band-width, or, putting the matter another way, the overall variations in the mean frequency which can be tolerated (considering the transmitter and receiver together) should be of the same order of magnitude as the variations of frequency produced by the modulation.

It should, however, be added that in some cases the maximum receiver sensitivity is a very unimportant factor. Under these conditions the principle stated above does not apply, and may, indeed, be misleading.

To clarify the discussion, it may be well at this point to remind the reader that the ultimate sensitivity of a receiver is a function of the noise at the *input* stages and is not a function of the noise *output*. For example, in comparing wide-band F.M. with A.M., the *output* signal/noise ratio of the former is large compared with that of the latter for all values of the *input* signal/noise ratio greater than a certain "threshold" value. But the "ultimate" sensitivity of the A.M. receiver is always greater than that of a corresponding wide-band F.M. receiver, since the former always gives a superior output signal/noise ratio for small values of the input signal. It is this ultimate sensitivity which is the criterion in the frequency control principle enunciated above.

There is another aspect of frequency control which is becoming more important as new modulation techniques develop. The frequency determining elements in transmitters and receivers are subject to two easily distinguishable forms of variation—

(*a*) long term drift which is basically aperiodic, although it may vary slowly about a mean value, and

(*b*) short term random or periodic variations, the fundamental frequency of which is in the audio or video region of the spectrum.

Variations of type (*b*) are of small importance in a simple A.M. system, but may seriously affect the performance of an F.M. or P.P.M. system. A particular form of the variations of type (*b*) is known as "frequency scintillation." To exemplify the effect of such short-term changes, consider an F.M. system in which the frequency determining element (either in transmitter or receiver) is a crystal oscillating at a frequency of 1 Mc/s. Let the carrier mean frequency be 100 Mc/s and the deviation for 100% modulation be 30 kc/s. Suppose that the effect of ripple in the power supply to the crystal oscillator is to produce a frequency modulation of one part in 10^5. Then the hum frequency modulation at the crystal would be 10 c/s, and at the carrier frequency would be 1 kc/s. In this case the 100% signal/hum ratio at the transmitter or receiver would be 30 db

(10 log 30^2). This a very severe limitation. A similar example might be concocted for P.P.M., since the mean epoch of the leading edge of the pulse is determined by the quality of the frequency control in the pulse circuits. It should be noted, however, that this form of frequency variation noise can always be overcome by increasing the modulation band-width. If, in the example given above, the deviation had been 300 kc/s, the hum level would have been 50 db below the signal. Fortunately, in most practical systems variations of type (*b*) are small compared with variations of type (*a*). It is not difficult to achieve a short-term frequency stability of one part in 10^9 in a crystal oscillator, whereas it is not at all easy to maintain the mean value of the frequency constant to one part in 10^7.

6.2. **Control by Piezo-electric Crystal.** By far the most popular means of determining the carrier frequency of a transmitter or the heterodyning frequency of a receiver is a crystal-controlled oscillator, the crystal almost universally used being made of quartz. This oscillator can be used in a large variety of ways, and modern communication practice has established several techniques which give great promise when applied to the U.H.F. region. The manufacture of crystals has undergone many major changes in the course of the last few years, as may be seen from Plate III, where some new crystal units have been grouped together. Plate III (*a*) shows a modern crystal, operating at a frequency of a few Mc/s, in which the dimensions have been radically scaled down. Plate III (*b*) is a new type of control component, made at the Marconi Research Laboratories in which several crystals are included in one valve envelope on a B7G base, while Plate III (*c*) shows a similar unit, employing the new overtone crystal which operates in the manner of an orthodox oscillator at frequencies up to 60 Mc/s. Plate III (*d*) is an example of modern American design, again using the so-called "modal" or "overtone" crystal, while Plate III (*e*) is a standard frequency bar as used in quartz crystal clocks.

It will be obvious from the above descriptions that one tendency in modern British crystal unit manufacture is to operate the crystal *in vacuo* and to use plated electrodes. The advantages of these techniques are considerable in relation to accuracy of setting, absence from ageing, and improved activity.

6.2.1. *Direct Crystal Oscillator Control.* The most common form of crystal control is that in which the only self-maintained oscillator is the one containing the crystal as the frequency-determining element. The carrier frequency is derived from the crystal by simple multiplication, division, addition or subtraction, the power of the oscillation being amplified wherever necessary. This type of control in the form of a multiplier chain has been employed experimentally at frequencies up to 3000 Mc/s both in U.S.A. and in Britain. In U.S.A. a klystron was used in the final stage, while in Britain triodes only were employed. As may be seen from Chapter III, the anode efficiency of micro-wave triode or klystron multipliers is very low, but it is sufficiently high to be a practical and economic proposition in some cases.

The following example of what may be done in this way may be instructive. It describes a transmitter and receiver chain designed and used by the author and his collaborators in Admiralty Signal and Radar Establishment in 1945–6.

In an endeavour to obtain the maximum frequency stability in the communication band at 2100 Mc/s, it was decided to employ the most stable crystal oscillator which could be obtained in portable form at the time. This appeared to be a unit developed at the Radio Engineering Research Station of the G.P.O., and operating at the comparatively low frequency of 152 kc/s. The complete oscillator comprised—

(i) A quartz bar, mounted in an evacuated container, forming one arm of the bridge in a Meacham circuit.(1) The crystal container and bridge were accommodated in an oven, the thermostat of which controlled the crystal temperature to a few hundredths of a degree.

(ii) A three-stage amplifier, which, with the bridge, formed the self-sustained oscillator. The amplifier was operated from a stabilized power supply.

A diagram of the Meacham oscillator is shown in Figure 65. Its action is as follows—

The bridge is shown on the left-hand side of the diagram. The arm *AC* contains the crystal, and, if it is desirable, a series capacitance. Arms *CB* and *AD* are stable, non-inductive resistances of standard pattern. Arm *BD* is a small tungsten lamp. If the bridge is balanced, oscillations will not occur, but if the bridge is sufficiently unbalanced in one direction, the feed-back from the amplifier will maintain oscillation. As the amplitude of the latter builds up, the tungsten wire in *BD* rises in temperature and its resistance changes. Let the values of R_1 and R_2 be so chosen that this change in *BD* tends to bring the bridge towards the balance point. Then the amplitude of oscillation will be controlled by the bridge, and may be stabilized at any desired level. Now the frequency variations in the free valve oscillator are fundamentally due to the fact that the oscillation amplitude is determined by the non-linear parts of the valve characteristics; a free oscillator cannot operate Class A. Hence the valve oscillator is sensitive to supply voltage fluctuations, and changes of amplitude mean changes in frequency. But the bridge stabilized oscillator may be operated Class A, since the tungsten resistance is the amplitude-limiting device. The Meacham circuit is therefore inherently more stable than any other oscillator which depends for its amplitude-limitation upon a volt-ampere characteristic which itself depends upon supply voltage.

Meacham claimed for his first oscillator (*loc. cit.*) that the long-term stability was a few parts in 10^8. The author can vouch for the fact that in comparing two G.P.O. oscillators of this type the frequency difference was never found to vary by more than two parts

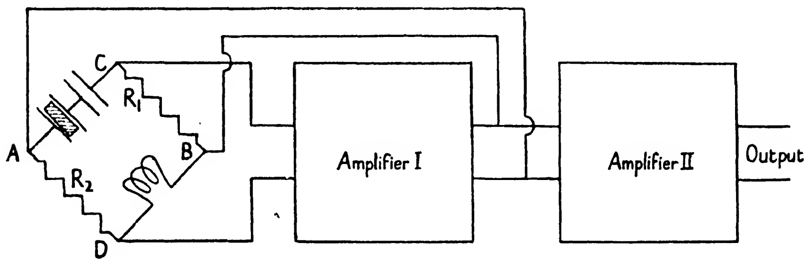


Fig. 65. Diagram of the Meacham bridge oscillator

in 100 millions. How much of this remarkable stability was due to the crystals and their temperature control, and how much was due to the circuit, cannot be estimated.

The output from the G.P.O. unit was about 100 mW at a frequency of 152 kc/s. Frequency doublers were successively applied to bring the carrier to 77·8 Mc/s, using double triode valves with separate cathodes and tuned circuits at each grid and anode. The output at this juncture was a few watts. A pair of CV53 valves, operated with balanced circuits and with their grids grounded converted to 233·4 Mc/s, when recourse was made to a concentric line circuit and a CV290* was employed as a trebler. Another CV290 with a concentric line cathode circuit and a wave-guide anode circuit gave an output power of about 150 mW at 2100 Mc/s. The overall multiplication factor was 13,824.

A second exactly similar chain was built, the crystal frequency being 152·4 kc/s. The frequency difference between the two after multiplication was approximately 5·53 Mc/s, and the difference frequency could therefore be compared with a standard crystal oscillator. By this means it was shown that the difference frequency was constant to within ± 20 c/s.† Now if x and y are the two crystal frequencies, and their difference frequency is n, all being measured in c/s, then

$$(x-y)1{\cdot}3824\,.\,10^4 = n$$

* This valve is now obsolete and is replaced by CV273.

† After additional smoothing had been added to the power supply, as described in the sequel.

and $\{(x + \delta x) - (y + \delta y)\}\ 1{\cdot}3824 \,.\, 10^4 = n + \delta n$

Therefore $$\frac{\delta x - \delta y}{x} = \frac{\delta n}{1{\cdot}3824 \,.\, 10^4 \,.\, x}$$

Let $\delta n = 40$ and $x = 1{\cdot}52 \,.\, 10^5$

Then $$\frac{\delta x - \delta y}{x} \doteqdot \frac{2}{10^8}$$

Thus the relative stability of the two oscillators was better than 2 parts in 10^8 over the period of observation. Readings were taken at random over a few weeks.

An Armstrong modulator was then interpolated between the crystal oscillator unit and the first frequency doubler, so that the

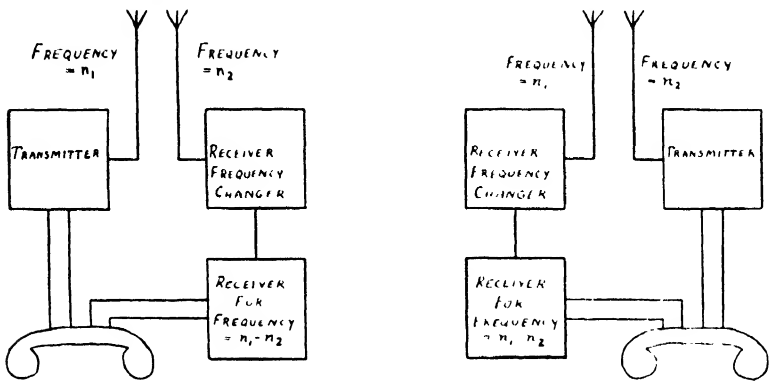


Fig. 66. Representing a two-way radio telephony link, employing "cross-band" working

chain could be frequency modulated. Each transmitter was linked via two aerials to a crystal frequency changer and a superheterodyne F.M./A.M. receiver, covering the band 4–8 Mc/s. By this means a "cross-band" radio-telephone link was formed, according to the layout of Figure 66.

The transmitting and receiving aerials at each terminal, consisting of half-wave dipoles, were mounted about 12 to 15 in. apart and parallel upon a wooden framework. Each transmitter, when unmodulated, then acted as the local oscillator for the reception of the other, the common frequency difference being fed at each end into the H.F. receiver. Normal two-way conversation was worked, care being taken to avoid "singing" due to feed-back at each hand-set.

As was mentioned in paragraph 6.1, there are two aspects of frequency control which must be considered quite separately, (*a*) the long-term stability of the system, and (*b*) the short-term or periodic variations in the audio region.

The system under examination had adequate long-term stability. To test the short-term stability, the transmitter was modulated, using 1000 c/s tone to give a deviation of 6 kc/s. The signal was then picked up in the receiver, the latter being close enough to give a strong carrier. In the first instance, before modifications were made, it was found that the power ratio

$$\frac{\text{signal} + \text{hum}}{\text{hum}}$$

was about 30 db, indicating that the swing at the carrier frequency, due to hum modulation, was 200 c/s. It was found that almost all of this emanated from the crystal oscillator, and was due to insufficient smoothing of the power supplies to the latter. At the crystal the frequency swing was 200/13,824 c/s, a fractional variation of about 1 part in 10^7. Remembering that two crystal oscillators contributed equally to this, the short-term variations in each must have been about 1 part in $2 . 10^7$.

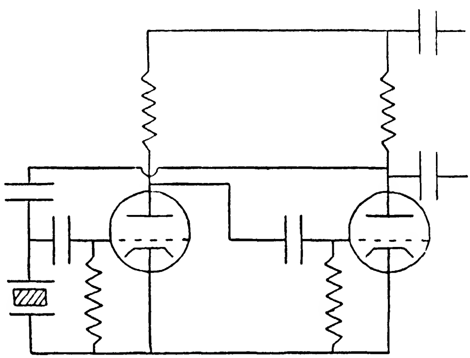


Fig. 67. A Franklin oscillator using a piezo-electric crystal in place of the tuned circuit

This is a very poor short-term stability. The variations were due to the sluggishness of the temperature compensating element in the bridge of the oscillator. When the smoothing was improved, there was a corresponding increase in the signal/noise ratio. The fact is, however, that the Meacham bridge circuit is not suitable for a narrow-band F.M. system of the type described. Much better results can be achieved using a simple oscillator of the type shown in Figure 67, and that without much loss in long-term stability.

This is a regenerative amplifier in which the crystal forms a high shunt impedance at its oscillatory frequency. Using such a device at a frequency of 4 Mc/s, the author has achieved a signal/hum ratio of 50 db without special smoothing or filtering, the carrier frequency being 3000 Mc/s and the deviation 6 kc/s, as before. This corresponds to a short-term stability better than 1 part in 10^8, divided between two oscillators.

The purpose of the analysis given above is to show how it is possible to check the stability of control at U.H.F. Very few practical systems would employ such a small deviation. If, however, a signal/hum ratio of 50 db can be achieved using a deviation of 6 kc/s, a practical system can be built, having a deviation of 600 kc/s and a signal/hum ratio of 90 db.

6.2.1.1. *Overtone Crystals.* One of the greatest problems in crystal controlled U.H.F. transmitters and receivers is to reduce the number

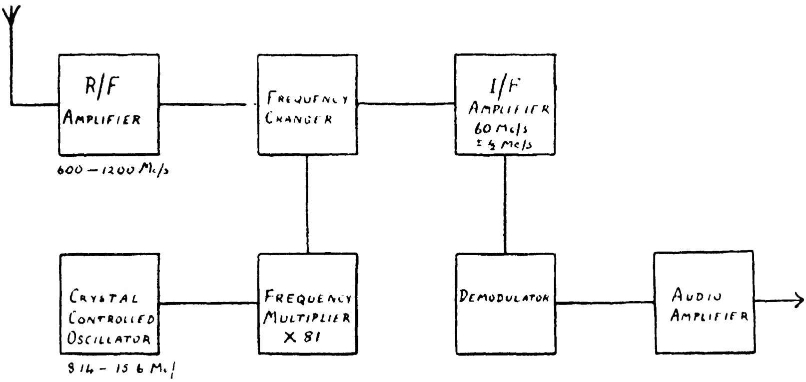


Fig. 68. A wide-band receiver for use at U.H.F., employing a crystal-controlled local oscillator

and amplitude of the unwanted harmonics of the crystal frequency. This is always a serious consideration, but in receivers intended to cover a wide frequency band it may become a major design factor. To exemplify the importance of these harmonics, consider a typical wide-band receiver such as is illustrated by Figure 68.

Suppose that this receiver is to cover the octave 600–1200 Mc/s, the frequency being determined by means of a local crystal oscillator. If standard BT-cut crystals were employed, a convenient range for the latter would be from 5–20 Mc/s. To obtain sufficient "image rejection" suppose that the intermediate amplifier is designed to operate at 60 Mc/s with a band-width appropriate to the stability of the frequency determining elements in the system. Let this band-width be $\pm \frac{1}{2}$ Mc/s, measured from half-power to half-power.

The crystal oscillators after suitable multiplication must therefore cover the band 540 to 1140 Mc/s, or, alternatively, 660 to 1260 Mc/s. Using a multiplication factor of 81, the range of crystals for this second range becomes 8·14 to 15·5 Mc/s. It will be noted that none of the directly generated harmonics of these crystals falls within the pass-band of the I.F. amplifier. Unless great care is taken, however, to avoid the generation and radiation of the 4^{th}, 5^{th}, 6^{th} and 7^{th}

harmonics, all crystals between 8·5 and 8·64, between 9·92 and 10·08, between 11·9 and 12·1, and between 14·9 and 15·1 Mc/s will swamp the signal. Hence there is a distinct possibility of the receiver being useless over the bands 628·5 to 640·1, 743·2 to 756·7, 903·9 to 920·1 and 1144·9 to 1165·1 Mc/s. This is a total loss of 61·6 Mc/s.

A recent development has assisted materially in solving this problem. Crystals can now be obtained in which the fundamental electrical frequency corresponds to a mechanical overtone of a particular shear mode of vibration. Such crystals are named "modal" in U.S.A. and "overtone" in Britain. The methods of manufacture are still in some degree secret, and it may be seen from Plate III, photographs (*c*) and (*d*) that the finished products differ considerably in the two countries. In general the American crystals show less tendency to oscillate at an unwanted frequency. The British crystals are, however, more active, and can be used in a greater variety of circuits. Clearly, certain steps can be taken during manufacture to encourage the wanted mechanical overtone at the expense of the fundamental, but even where this has not been deliberately done (as in the case of older crystals which were intended to be used on the mechanical fundamental), suitable circuit techniques can be adopted to suppress the lower electrical modes. The important factor is, of course, the resonant frequency of the anode circuit of the valve oscillator.

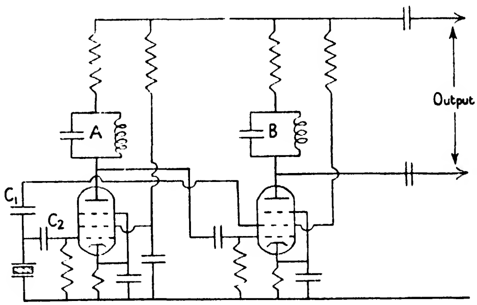


Fig. 69. A circuit for use with an overtone crystal as a local oscillator of a receiver. Circuit A is tuned to the desired mechanical overtone. Circuit B may be tuned to the same overtone or to a harmonic to give frequency multiplication

The most popular mechanical harmonic is the third, but crystals can be made to oscillate quite successfully on the 5^{th} and 7^{th} harmonics. This new technique empowers the communication engineer to use crystal control in such a way that the lowest oscillator frequency present in a receiver is higher than the intermediate frequency at which the major amplification occurs. In this manner the troubles described above are automatically eliminated. Even where the lowest oscillator frequency is not above the intermediate frequency, the use of "overtone" crystals reduces the local interference. A typical circuit for use with an "overtone" crystal is shown in Figure 69. As may be noted, this is a modification of the Franklin oscillator (Figure 67), the crystal, operating in a parallel resonance, replacing the tuned circuit. Using a pentode as the first valve, the pulling effect of the anode circuit is not very marked, and in the second valve the anode circuit may, if it is so desired, be tuned to a harmonic.

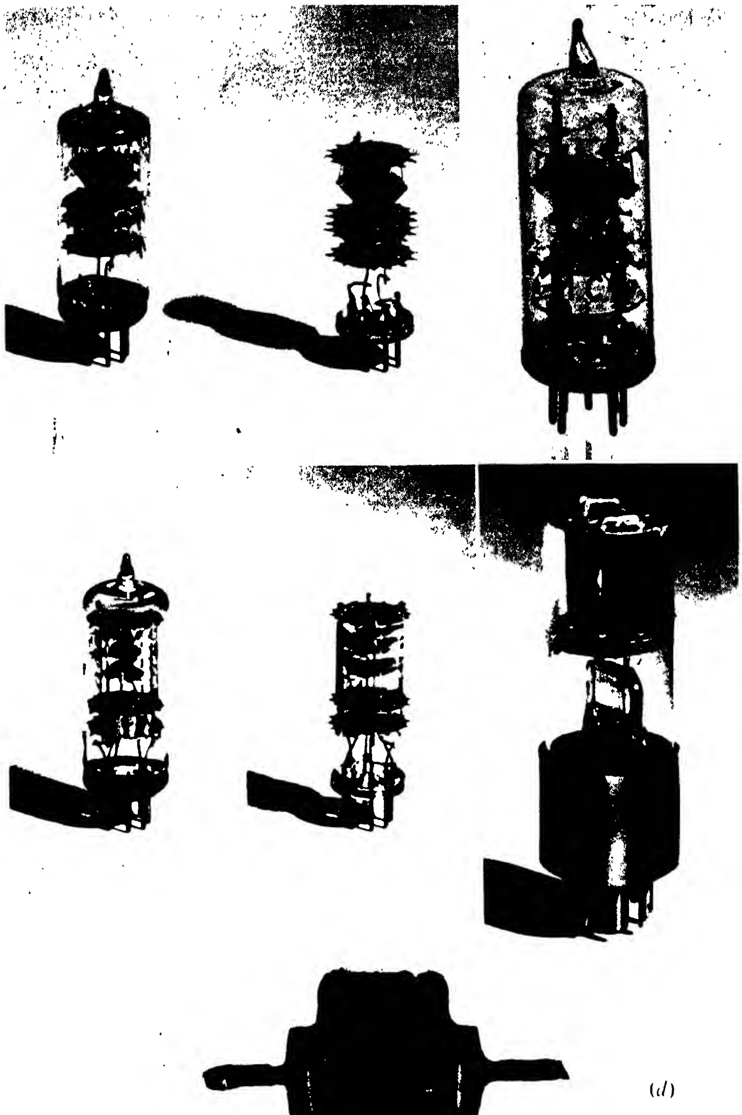
Typical operation for a crystal, the mechanical fundamental of which is at 10 Mc/s, would be as follows: The fundamental electrical oscillation at about 30 Mc/s would be encouraged by tuning circuit A to this frequency. The feed-back from the screen-grid of the second valve, which operates as a triode-tetrode, would maintain the oscillation. The condensers C_1 and C_2 should be as small as possible—of the order of 1 *p*F each, including strays. Circuit B might be tuned to 90 Mc/s, giving a reasonable output on the third harmonic.

Another popular circuit for use with "overtone" crystals is one described by Butler[(2)] in 1946. This uses the crystal in the series-resonant mode, and again employs two valves. Of the two circuits the first is the more stable, but is more difficult to operate at the highest frequencies.

Referring back to Figure 68, if the local oscillator employs overtone crystals within the range 24·4 to 46·6 Mc/s with a multiplication factor of 27, only those crystals in the range 29·75 to 30·25 Mc/s can obliterate signals. This corresponds after multiplication to a signal-frequency range of 13·8 Mc/s, less than one-quarter of that obtained using the older type of crystal. By reducing the intermediate frequency to 40 Mc/s all possibility of obliteration can be eliminated and in this case crystals between 41·25 and 78·75 Mc/s would be used.

The use of overtone crystals has another advantage in many cases. The degree of frequency multiplication employed in a transmitter or receiver determines in some measure the number of spurious responses of which various circuits are capable. In one sense the obliteration effect just described is an example of this, but it is a very special case and has therefore been treated separately. In the more general case there is always the possibility of a transmitter radiating power on an unwanted frequency, or of a receiver responding, however weakly, to an unwanted signal. The use of overtone crystals diminishes this danger, and simplifies the design of frequency multipliers. There is also a saving in valves which in certain circumstances can be important.

The general performance of overtone crystals is satisfactory compared with their fundamental counterparts. Cutting and grinding do not appear to offer any insuperable difficulties and their activity, though much lower than that of BT-cut crystals operating at one-



(a) (c) (b) (e (d)

MODERN QUARTZ CRYSTAL UNITS

(a) Modern crystal unit on B7G base.
(b) Three crystals in the same envelope.
(c) An overtone crystal unit.
(d) American overtone crystal unit.

third of their frequency, is adequate for most purposes. Temperature coefficients of the order of one part in a million per Centigrade degree over a range of 10 degrees have been achieved. Frequency drift due to ageing is not greater when working on the mechanical overtone. In fact, the only real difficulty is that the oscillator in which an overtone crystal is used must contain a circuit the resonant frequency of which is in the neighbourhood of the overtone frequency. This places a theoretical limitation on the stability of the

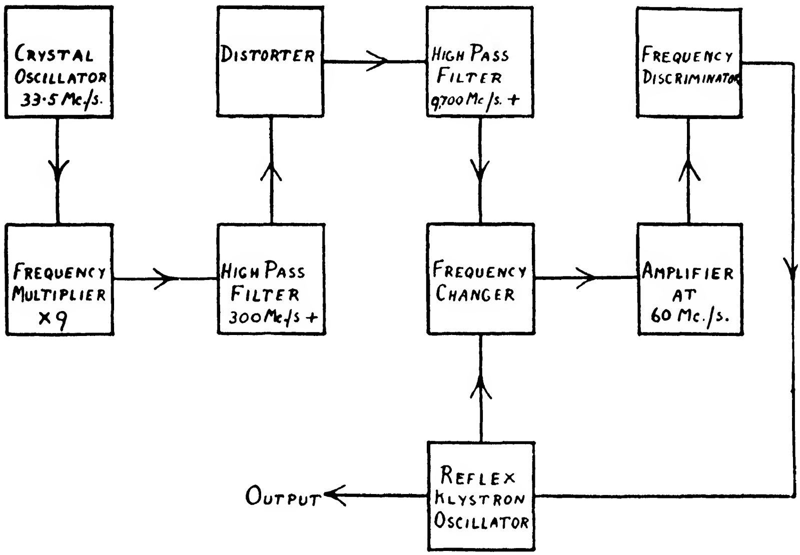


Fig. 70. Block diagram of a unit giving indirect crystal control of the frequency of a reflex oscillator and applicable at frequencies up to 10,000 Mc/s

oscillator which does not exist with such well-known circuits as those due to Pierce and Franklin.

6.2.2. *Indirect Crystal Control.* At U.H.F. it is sometimes inconvenient for various reasons to control the frequency by direct multiplication. This has produced a variety of new techniques of indirect control, some of which will be described in this and succeeding paragraphs. The most obvious method is exhibited by Figure 70 in which frequencies have been inserted to serve as an example.

The alternative to a bulky chain is to start with an overtone crystal at about 33·5 Mc/s and to multiply twice by three to obtain an output of the order of 1 watt at 301·5 Mc/s. This is then filtered by a high pass circuit and distorted by a rectifier, giving a series of harmonics of 301·5 Mc/s. Those in the vicinity of 10,000 Mc/s are filtered through a short section of the appropriate wave-guide to a silicon crystal frequency changer where they are mixed with a small fraction of the output from the reflex klystron oscillator. The latter is tuned until its frequency is 60 Mc/s away from that of the chosen harmonic. The output from the frequency changer is then amplified and applied to a stable discriminator. The D.C. output from the latter is fed back to the reflector of the klystron. This automatic frequency control, as described in Chapter III, maintains the frequency of the oscillator at an almost constant value. A simple

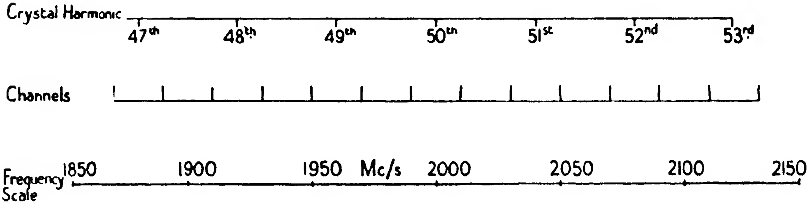


Fig. 71. To show how the harmonics of a crystal oscillator may be used to control the frequencies of many channels at U.H.F.

wave-meter is sufficient to inform the operator if the oscillator frequency is above or below the desired crystal harmonic, and to ensure that the correct harmonic is being used. This removes any ambiguity from the system.

It will be observed that this indirect control implies two frequency determining elements; the crystal and the 60 Mc/s discriminator. The centre-frequency of the latter may vary by 60 kc/s, since a stability of one part in a thousand is as high as may be expected. But this only contributes a variation of 60 kc/s to the klystron frequency, and so, if the crystal oscillator were perfect, the stability of the klystron would be of the order of six parts in a million. It can therefore be seen that although the discriminator is a frequency determining element in the system, the part it plays is very small and it is usually possible to neglect its variations.

There are some obvious extensions of this control mechanism. One is to locate the oscillator on any one of a number of equally spaced channels by using adjacent crystal harmonics. For example, considering the 2000 Mc/s region of the spectrum, if the harmonics of a 40 Mc/s oscillator are employed with an I.F. amplifier and discriminator centred on 10 Mc/s, then channels spaced 20 Mc/s apart can be stabilized. This is done by using each harmonic twice as the heterodyning oscillator. The channels are shown diagrammatically in Figure 71.

Another extension which is not so obvious, perhaps, is to employ two discriminators in series to improve the stabilization. By this means the frequency of a reflex klystron operating at 10,000 Mc/s can be maintained at as constant a value (relatively speaking) as the crystal oscillator forming the chief frequency determining element. A circuit used by Borg[(3)] is shown in Figure 72.

It was experimentally verified that this device constrained the klystron to oscillate at a frequency equal to that of the desired crystal harmonic plus or minus less than 100 c/s.

6.2.3. *Partial Crystal Control.* Where it is necessary to cover a frequency band without gaps, the method known as partial crystal

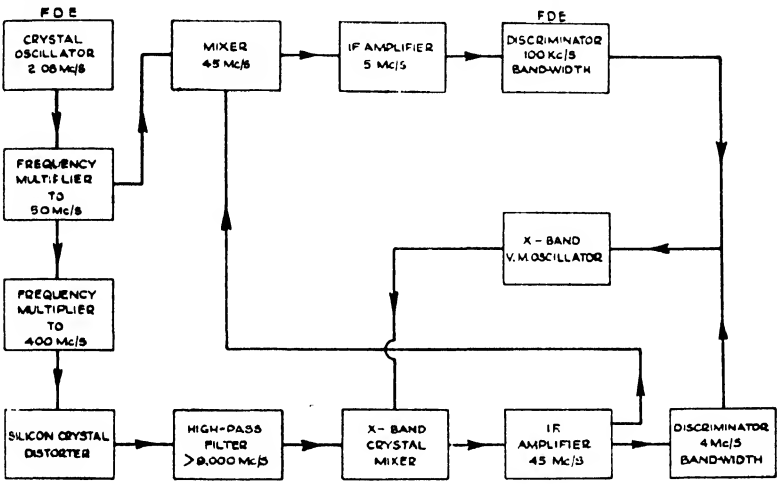


Fig. 72. A block diagram of a unit employing a coarse and a fine discriminator for the exact location of a frequency in the neighbourhood of 10,000 Mc/s

control is sometimes valuable. The principle involved is the addition (or subtraction) of two frequencies, one fixed and due to a crystal-controlled oscillator, the other variable and due to an L-C oscillator. The principle is illustrated by Figure 73 which represents the radio-frequency chain of a transmitter designed to operate at any frequency between 144 and 168 Mc/s. In this case the two frequencies have been added before multiplication. There are many problems associated with frequency composition techniques of which this is an example. The stability of the carrier frequency depends upon the stabilities of both the crystal and the variable frequency oscillator, and if the latter is to have the same net effect as the former, its highest frequency must be very much smaller than the crystal frequency. An even more important problem is that of eliminating unwanted frequencies. These are inevitably generated in the various non-linear devices in the chain and must be suppressed by some simple means. These "spurious responses" are discussed in detail by Hupert[4] who suggests methods of reducing their more detrimental effects. One of the most troublesome effects encountered in a frequency composition scheme is the phase-modulation of a wanted frequency by a near-by unwanted frequency of much smaller amplitude. After frequency multiplication this phase-modulation

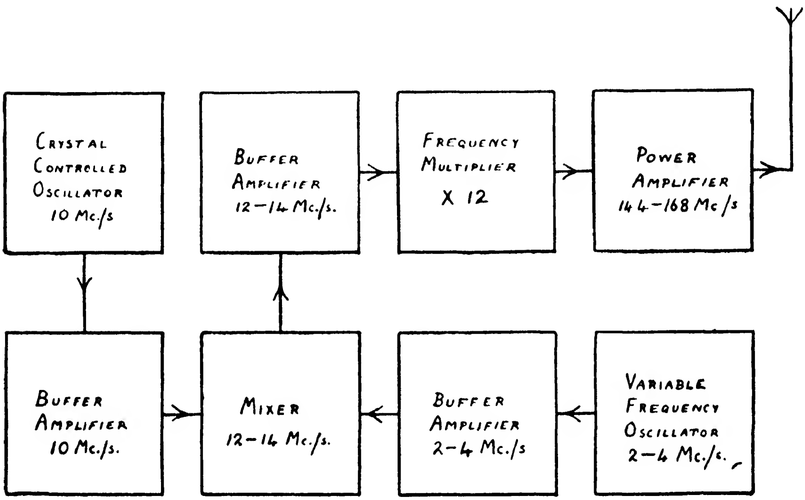


Fig. 73. A block diagram illustrating the principle of partial crystal control of frequency

can be so increased that it will effectively interfere with the modulation components.

In most applications spurious responses should be reduced by appropriate techniques to a level at least 60 db below the wanted frequency. In certain cases it is necessary to work to a level of − 80 db, and this may be very difficult indeed.

6.2.4. *Frequency Composition.* The principle exemplified by paragraph 6.2.3 can be extended in several directions. One important extension is to the problem of producing and defining a large number of channels in a given band. If these channels are spaced equally on the frequency scale, they may be obtained by two or more frequency compositions, using a sub-multiple of the frequency difference between the channels as a fundamental unit in the composition. A simplified version of such a composition scheme (or frequency synthesizer) suitable for the definition of 1000 channels is shown in Figure 74. Since the fundamental unit here chosen is 1 kc/s, the scheme would only apply directly to a system in which the channels were n kc/s apart and n was a simple multiple, such as 2, 2 × 2,

2×3, 3×3, etc.; $n = 10$ would not be suitable, since multiplication by 10 is not a practical possibility, using conventional valves.

This is an example of a decade system in which each crystal oscillator is used on 10 harmonics. It is therefore suitable for linking to a simple uniselector. All three crystal frequencies could be derived by frequency division from a single crystal, if such were

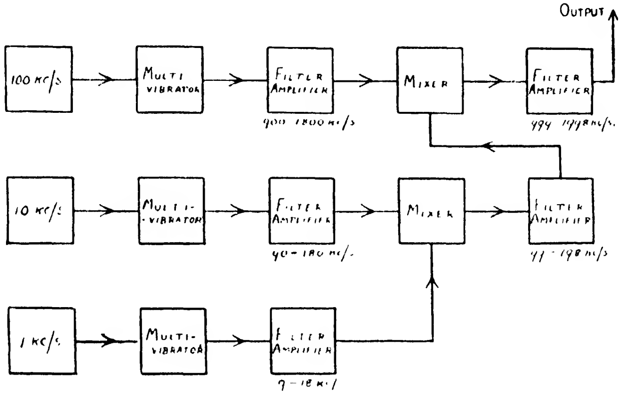


Fig. 74. A block diagram illustrating the principle of composition

desirable. The problem of spurious responses is very serious indeed in such a unit, since coincident frequencies are produced by all three chains, and phase changes in each may therefore produce phase- or frequency-modulation of a frequency in another. However, where elaboration of circuits is possible, and particularly where adequate filter amplifiers can be provided, such a decade composition unit may be made to work very well. It will be noted that it contains seven independently tuned sub-units. In general each sub-unit contains two ganged circuits.

6.2.5. *The Choice of a Crystal Controlled Oscillator.* In the search for ever-greater stability at U.H.F. the engineer finds himself forced to consider refinements of crystal oscillator technique which are of little importance at lower frequencies. Although it is not proposed to enter here into a detailed study of the various circuits, it is worth while to bring into focus the different design factors which determine the engineer's choice.

In cases where a U.H.F. equipment is required in relatively small numbers, the choice of an oscillator is a comparatively simple matter, for it is feasible to be meticulous with regard to tolerances. This applies equally to the cutting and grinding of crystals and to the specification of the other components such as resistances and capacitances. In cases where the equipment is required in large numbers, the designer is forced to use components made to much wider tolerances, or, at least, it should be his aim to do so. He must therefore decide where he is going to lay the emphasis for frequency determination, and he must then try to make his circuit so that the frequency is as independent as possible of variations in other components. In general, it is good practice to use the crystal itself as the frequency determining element, and to build the oscillator round this in such a manner that all other components are unimportant. This may seem rather obvious, but in the past such a dictum has not always been obeyed. It must be decided whether the crystal is to be cut for series or parallel resonance and what degree of amplification will be used to obtain the desired output. A compromise must always be reached between the isolation of the crystal which is desirable for frequency stability, and the crystal coupling which is necessary to obtain the desired output in a certain number of stages. For U.H.F. work it is generally necessary to employ two valves as an oscillator and amplifier, or as an oscillator-amplifier, before attempting to use any power at the crystal frequency. Examples of such circuits are shown in Figures 65, 67 and 69. This is good practice even where it is desirable that the equipment should be small and simple, for the advantage gained is a fundamental one. By isolating the crystal so that it is the major frequency determining element, the tolerances on other components may be relaxed; and by improving the frequency stability, the receiver band-width can be narrowed with corresponding improvements in sensitivity and selectivity.

6.3. **Control by Conductors.** Although crystals have been made to oscillate on an electrical fundamental as high as 150 Mc/s, the engineering use of crystal frequencies much above 50 Mc/s is some way in the future. Hence crystal control of frequency at the present time inevitably implies multiplication with its attendant difficulties. In the attempt to avoid these difficulties, metallic circuits have been employed at U.H.F. as frequency determining elements. In the micro-wave region, in particular, this technique has met with a measure of success. The theoretical basis for the use of resonant lines or cavities is that very high Q's can be obtained by means of these devices, so that rapid phase or amplitude variations take place in the vicinity of the resonant frequency. A subsidiary advantage which may be gained from the use of such circuits is that they may be tuned over a comparatively wide band, thus giving flexibility of frequency definition. Commercial equipment exists in U.S.A., employing "long line" frequency stabilization. This has proved to be very successful, particularly where space is not a consideration.

Whether the principle is equally applicable at higher frequencies to more compact equipment remains to be tested. In the succeeding paragraphs some of the more outstanding features of such conductors will be described.

6.3.1. *U.H.F. Circuits of High Frequency Stability.* Imagine a circuit of high stability such as a quarter-wave concentric line or a resonant cavity connected as shown in Figure 75 so that a small fraction of the power from an oscillator is fed into it at A and the rectified output establishes a potential difference at B. This potential

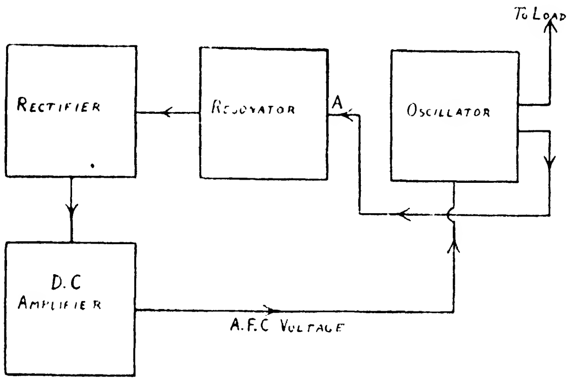


Fig. 75. A block diagram of a simple frequency control unit employing a resonant cavity of high Q as the frequency determining element

difference may then be amplified and fed back as a compensating voltage for the automatic frequency control of the oscillator. An improvement can be effected by using two resonators adjusted to be equidistant from, but on opposite sides of, the desired frequency. Such a balanced arrangement eliminates the effect of changes of load, since variations of the input to the discriminator produce very much smaller variations of the output. The balanced system has the corresponding disadvantage that it involves two resonators, each of which must be adjusted, and each of which may go out of order.

Considering such a balanced discriminator, let the D.C. amplification be such that a change of 1 microampere in a standing current of 100 microamperes can be amplified to produce a systematic change of frequency in the oscillator. If both resonators have equal Q factors, then 1 μA will correspond to a change of frequency of the order of $n/100Q$, where n is the desired frequency. This is assuming that the two resonators are adjusted so that the frequency n is about half maximum amplitude in both; that the resonators are adjusted, in fact, as shown in Figure 76.

A reasonable value for Q in the case of a quarter-wave concentric line is 1000, giving a frequency discrimination of the order of $n/10^5$, At micro-waves, where a Q of the order of 20,000 can be obtained. the discrimination would be $n/2 \cdot 10^6$. This is of the same order of magnitude as can be achieved by crystal control, providing that the resonance frequencies of the resonators themselves are stable to this degree.

Consider now the temperature coefficient of change of frequency. The wave-length λ of the resonator is proportional to its linear

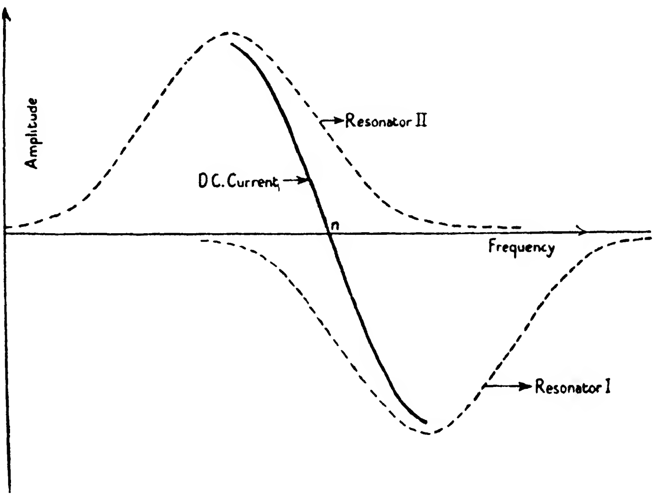


Fig. 76. Diagram illustrating the relative adjustment on the frequency scale of two resonators forming the elements of a balanced discriminator for use in frequency control

dimensions, and each of the latter is proportional to $(1 + \alpha t)$, where α is the coefficient of increase of length with temperature, and t is the temperature, measured from some arbitrary zero.

Therefore $$\lambda = K(1 + \alpha t) = \frac{c}{n}$$

and $$d\lambda = K\alpha dt,$$

and $$d\lambda = -\frac{c}{n^2} dn$$

Therefore $$dn = n\alpha dt$$

If this quantity is to be small compared with the discrimination provided by the circuit, then for micro-waves at least,

$$\alpha dt \leqslant \frac{1}{2 \cdot 10^6} \text{ or, let us say, } \alpha dt = \frac{1}{2 \cdot 10^7}$$

If a copper resonator were used, the temperature would therefore have to be controlled to $\pm$ 1/1000th of a degree Centigrade.

Clearly, for a resonator to act as a satisfactory standard of frequency in the micro-wave region, not only must its linear expansion be very small, but its temperature must be controlled within fine limits. Many laboratory experiments have been made to find such a standard. An obvious technique is to employ a resonator consisting of invar or even silver-plated quartz. By this means the coefficient of expansion may be reduced to a few parts in 10^6. A better method is to use two dissimilar metals to derive a temperature-compensated system, and to operate this at the optimum temperature, where the coefficient of expansion is changing from positive to negative or vice versa. Under these circumstances a reasonably wide temperature variation can be tolerated. Two metals which have been used in this way are copper and molybdenum. Having overcome the effects of temperature variation, the resonator cannot yet be regarded as a satisfactory standard. Ageing effects due to the creep of the metal are significant, as are the variations in properties from one sample of the material to another. On the whole, therefore, it may be said that a metallic circuit at U.H.F. offers certain possibilities for short-term automatic frequency control, but it is unlikely to compete with the more stable and well-defined piece of quartz for absolute control. Despite this, the fact that the resonator is tuneable gives it a genuine utility in certain applications.

6.3.4. *Combination of Resonator and Crystal Oscillator.* Where a number of channels are required and where, therefore, a resonator control is desirable, the worst effects of long-term instability mentioned above can be much diminished by the use of a crystal calibrator. This is not a positive crystal control in the usual sense of the term, but the existence of the necessary number of harmonics of the crystal frequency allows the resonator to be calibrated from time to time. Moreover, since the crystal calibrator is only required at intervals, the power drain is not excessive. Such a system is a good practical compromise for many U.H.F. applications.

6.4. **Control by Molecular Absorption.** Atomic theory is sufficiently well authenticated with regard to radiation phenomena to allow the engineer to use with confidence formulae defining the specific absorption frequencies of certain molecules and atoms. In the millimetric wave region several gas molecules exhibit pronounced line absorption, notably oxygen and ammonia, and the absorption frequencies may be accurately calculated from wave-mechanical theory. Hence, if an electronic oscillator can be produced to work over the required band, its frequency can be determined in terms of atomic constants. This suggests an entirely new approach to the problem of frequency definition at micro-wavelengths.

Absorption "lines" have a frequency distribution which is a function both of the absolute temperature and of the pressure. The broadening of spectral lines due to these causes has been known for many years. If, therefore, a tube containing ammonia gas at a very low pressure and temperature is employed as an attenuator, the "line" is fairly sharp, and a rapid variation of the attenuation will be obtained in the vicinity of the absorption frequency. This variation can then be used to control the frequency of the valve generator at the other end of the attenuator by feeding back the rectified output in the correct manner. This method of frequency control, first suggested by members of the Bell Telephone Laboratories, is now being examined both in U.S.A. and in Britain. It will be some time, however, before practical control circuits are developed which give the improved performance relative to a quartz crystal which theory predicts.

REFERENCES

(1) *B.S.T.J.*, p. 574 (1938).
(2) *Wireless Engineer*, 23, 273, p. 157 (1946).
(3) *J.I.E.E.*, Pt. IIIA, 11, p. 107 (1947).
(4) *J.I.E.E.*, Pt. IIIA, 12, p. 405 (1947).

CHAPTER VII

COMMUNICATION SYSTEMS

7.1. Modulation. In the chapter specifically concerned with modulation techniques little attention was paid to the effect of the latter on the system. It is the purpose of this section to discuss the advantages and disadvantages accruing from the use of each technique in relation to performance; no account will be taken of the complexity of the equipment.

It has already been pointed out in Chapters IV and VI that at U.H.F., where receiver noise is the limiting factor, the smallest signal which can be detected in a given system is a function of the effective band-width of the modulated signal stages of the receiver and therefore of the relative frequency stability of the transmitted carrier and the receiver's heterodyning oscillator or oscillators. This relative stability is, therefore, a most important parameter in system performance.

If a communication system is to be designed to give the maximum response to the minimum signal, suppressed carrier single side-band amplitude modulation is theoretically supreme, assuming that the required frequency stability is attainable. Fortunately for the designer, it is very seldom indeed that "the maximum response to the minimum signal" is of paramount importance.

On the other hand what is very often required is the minimum input signal for a certain value of the output signal/noise ratio. This is perhaps difficult to understand, but it can be made clear by reference to Figure 77 below.

Imagine a system in which the output signal/noise ratio from the receiver is a linear function of the input signal. This is represented by the broken line. To a very rough approximation a simple A.M. system is characterized by this type of response. Now suppose that it is found that the minimum output signal/noise ratio of any value to the operator is represented by the level marked AA′. Then the input signals to the left of point B are useless for communication, and B may be termed the "threshold" value for the system. Imagine another system represented by the curve KLM. In this the threshold value is C, and, since C is to the right of B, the second system is poorer in performance than the first. Similarly, a third system, represented by NOP, and having a threshold value of D, is superior. If, then, the performance criterion of the system is as stated above —the minimum input signal for a given output signal/noise ratio—

it is not necessarily true that the system should exhibit maximum sensitivity for the minimum signal. Once this is properly appreciated much confusion with regard to the assessment of various systems can be avoided. In voice communication it is questionable if values of the *maximum* output signal/noise ratio less than 5 db are of any interest. Harvard Psycho-Acoustical Laboratory in an exhaustive series of tests has reported that speech can be intelligible

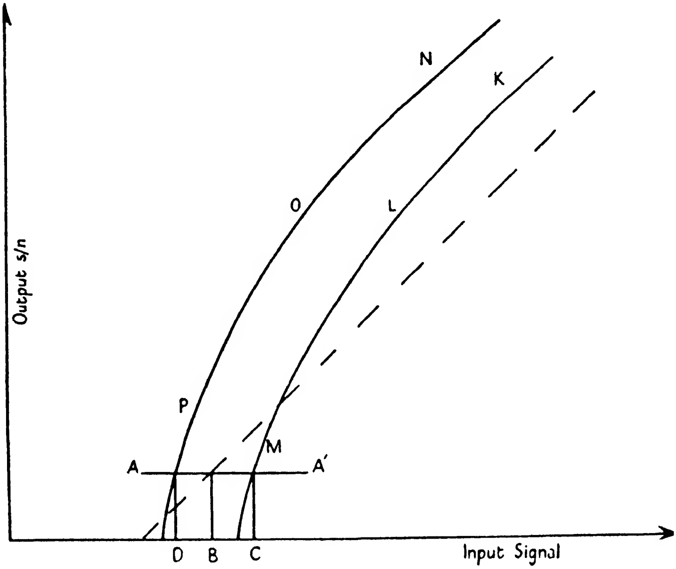


Fig. 77. Typical curves showing the variation of the output signal/noise ratio with the input signal to illustrate the importance of the conception of a minimum usable signal

when the ratio of signal/noise is less than unity, meaning that a certain small fraction of the intelligence can be appreciated. The result should be ignored for practical work. It is as if a typewriter manufacturer announced that his machine would still yield a modicum of intelligence after passing it through a mincing machine. However true, it is a little irrelevant. Indeed, there is a widely accepted international standard for the minimum signal/noise ratio in a circuit carrying speech of tolerable intelligibility. This is 15 db measured with 100% tone modulation, or, assuming that the average speech level corresponds to 33⅓% modulation, the standard corresponds to an *average* of 5 db. This figure is to a very large extent independent of the system of modulation, whether the latter be amplitude, frequency, code or pulse. Consequently it forms an excellent basis upon which to assess the value of a system. The independence mentioned above is not obvious *a priori*. It has only been discovered by experiment and it means that the type of distortion introduced by each modulation technique has the same effect on intelligibility at the same noise level.*

7.1.1. *Amplitude Modulation.* This, the best-known and most thoroughly investigated system, forms a convenient standard with which to compare all others. It is necessary, however, to define the modulator and demodulator more closely, for there are many

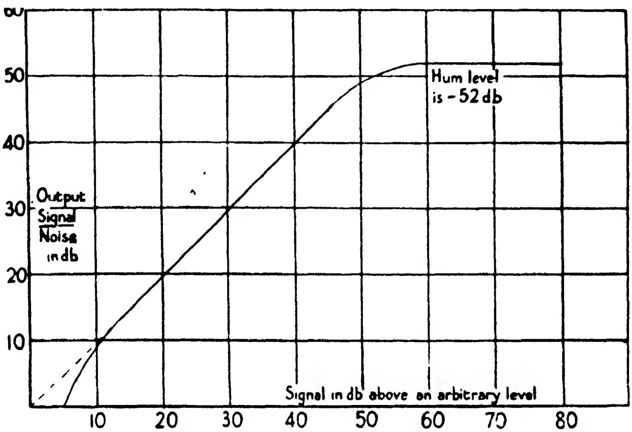


Fig. 78. Idealized curve of output signal/noise ratio as a function of input signal for an A.M. system

system modifications of amplitude modulation. Assume, then, that the standard A.M. system is one in which there is no pre-emphasis or de-emphasis, no automatic gain control, no form of noise limiter, and an audio pass-band limited to 300–3000 c/s. Any other standard would give the same system comparison, but this is simple and well understood.

The effects of modulation characteristics upon the system are exhibited in a compact form by a graph showing the mode of variation of the output signal/noise in terms of the input signal. This may be termed the "response curve" of the system. It should be noted that it is as much a function of the transmitter as it is of the receiver. Amplitude modulation, as previously mentioned, is characterized by a roughly linear response curve, so long as there are

* It is now widely recognized that at low signal/noise levels the effect of an increase in distortion on the intelligibility is very much like an increase in the noise. Indeed, the threshold of 5 db, quoted above, is 10 db less than the international standard, 15 db, only because more recently engineered systems have suffered less from distortion. Reducing the harmonic distortion from 10 per cent to 2 per cent reduces the threshold by about 10 db when the background noise is thermal.

no additions such as A.G.C., etc. The non-linearity which exists is due to the fact that the intercombination noise terms become of greater significance at the lower signal levels. Hence, the non-linearity depends entirely upon the band-width of the receiver stages handling the modulated signal. If this band-width is large, the response curve shows considerable curvature; if the band-width is so narrow that only the modulation can be accepted, the response curve is virtually linear. A slightly idealized A.M. curve is shown

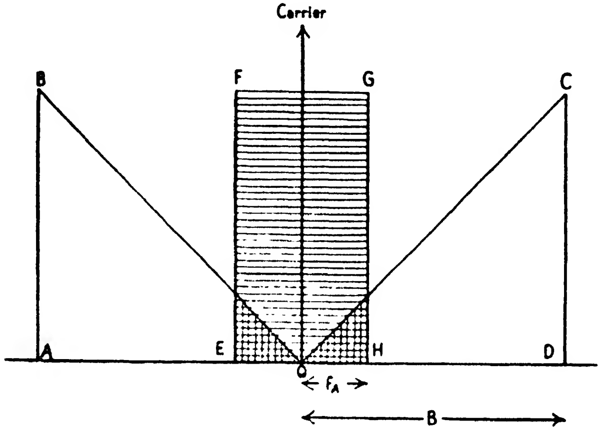


Fig. 79. Diagram showing the "triangulation of noise" in an F.M. system

in Figure 78, where the slope at small signal/noise ratios changes quite rapidly to a slope characteristic of the audio pass-band. This remains fairly constant over a range of signal strengths depending upon the ability of the receiver amplifiers to operate without limiting. Ultimately the audio hum level determines the maximum signal/noise.

7.1.2. *Frequency Modulation.* For large signal/noise ratios the improvement of an F.M. over a corresponding A.M. receiver is $20 \log \sqrt{3}m$, where m is the modulation index commonly called the deviation ratio. This improvement is sometimes said to be due to the "triangulation of noise." Since there might be some doubt as to what is meant by a "corresponding" receiver, the above formula will be proved, and in the proof the definition will emerge.

In Figure 79 let the rectangle ABCD indicate the limits of the pass-band of the intermediate frequency amplifier of a receiver, the band-width being 2OD = 2B. Let EFGH represent the limits of the audio pass-band, where OH $= f_A$.

(*a*) Consider first an A.M. receiver of this type. Let the noise power available at the 2^{nd} detector be given by kB. The equivalent voltage is V, where $V^2/R_1 = kB$, and R_1 is a resistance. Then $V = \sqrt{kBR_1}$. Let the voltage conversion factor of the detector be θ_A. Then the resultant voltage is $\theta_A\sqrt{kBR_1}$ and the output power is $\theta_A{}^2kBR_1/R_2$. R_2 is another equivalent resistance. The filtering action of the audio circuits restricts the noise to a band $0 - f_A$. Hence the output audio power is $\theta_A{}^2kR_1f_A/R_2$.

Similarly, let the signal power available at the 2nd detector be NkB. Then the signal voltage is $\sqrt{NR_1kB}$ and the output signal power is $\theta_A{}^2NkR_1B/R_2$. The signal/noise ratio at the output is, therefore,

$$\frac{NB}{f_A}$$

This is only true when the ratio is large, for it is only under these circumstances that the intercombination tones in the noise spectrum may be neglected compared with the tones produced by the noise spectrum beating with the carrier.

(*b*) Consider next an F.M. receiver in which the discriminator response varies linearly with the frequency deviation. This is represented by the line OC of Figure 79. Again let the noise available at the detector be kB, corresponding to a noise voltage $= \sqrt{kBR_1'}$. This noise voltage is transformed by the discriminator into a voltage $\theta_F f\sqrt{kBR_1'}$, corresponding to a power P, where

$$P = \frac{1}{B}\int_0^{f_A} \frac{\theta_F{}^2f^2kBR_1'}{R_2'}\,df = \frac{f_A{}^3}{3}\,\frac{\theta_F{}^2kR_1'}{R_2'}$$

Here the mean square voltage has been taken, but the integration has been arrested at f_A to take account of the filtering action of the audio circuits.

The assumption behind this is that the noise variation with frequency and the noise rate of change of variation with frequency are both completely random.

Similarly, if the signal power is NkB, giving a voltage of $\sqrt{NkBR_1'}$, then the output voltage for a deviation f_D is $\theta_F f_D\sqrt{NkBR_1'}$ and the signal power output is $\theta_F{}^2f_D{}^2NkBR_1/R_2$. This makes the F.M. signal/noise at the output

$$\frac{3f_D{}^2}{f_A{}^2}\,\frac{NB}{f_A}$$

Hence, if equal signals are applied in the two cases, the receivers being "corresponding," the signal/noise improvement is $3f_D{}^2/f_A{}^2 = 3m^2$, or in decibels $20\log\sqrt{3}m$. Clearly, the definition of a "corresponding" receiver is a receiver in which the noise factor is the same. The I.F. band-width is not a factor in either case, and the R.F. band-width only affects the result in so far as it may modify the noise factor. Returning to Figure 79, the F.M. noise is represented by the two cross-hatched triangles, and the A.M. noise by the rectangle EHGF.

Typical response curves for an F.M. system are shown in Figure 80. This set has been drawn for an I.F. band-width of 100 kc/s, and an audio band-width of 3 kc/s. Curve A is for an A.M. receiver,

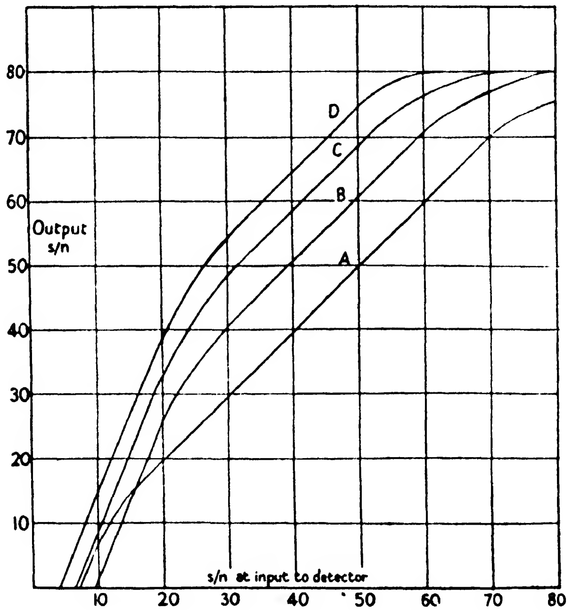


Fig. 80. Typical curves showing output signal/noise ratio as a function of input signal for F.M. systems. Curves *B*, *C*, *D* correspond to deviations of 6, 15 and 30 kc/s respectively. Curve *A* is for an A.M. system in which the receiver has the same noise factor and the same band-width

working under these conditions. Curves B, C and D are for an F.M. receiver, the deviations being 6, 15 and 30 kc/s respectively. There is a certain doubt regarding these curves. The so-called "threshold value" of the input signal/noise ratio is fairly well defined,(1) and the slope of the curve as the output signal/noise ratio approaches unity can also be estimated. For the former Crosby gives $10 \log 6{\cdot}25\ B/f_A$ on the db scale, where B is the band-width of the I.F. amplifier and f_A is the audio upper limit, and for the latter both experiment and theory suggest a value between 2 and 2·5 db/db. But the nature of the curve connecting the two lines is still a matter of guesswork. All that can be said is that the curves of Figure 80 have been experimentally checked and found to be correct within a few db in the dubious region.

7.1.3. *Pulse Modulation.* Noise in pulse modulation is discussed in some detail by Jelonek,[2] and from his results the response curves of Figure 81 have been constructed. As in the case of other modulation systems three data are required. First, the signal/noise

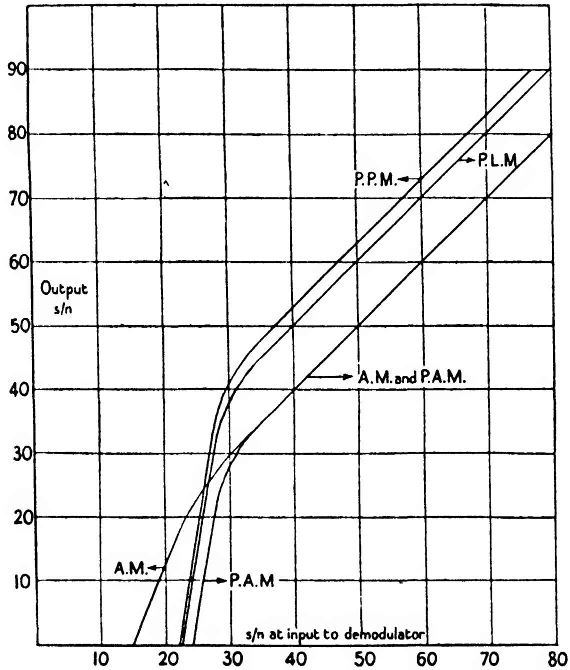


Fig. 81. Typical curves showing output signal/noise ratio as a function of input signal for A.M., P.A.M., P.L.M. and P.P.M. The conditions are detailed in the text

improvement over a standard A.M. system must be known. The figures given by Jelonek are quoted in Table VII below in terms of power (P). The corresponding figure in db is, of course, 10 log P. The figure for F.M. is given for comparison.

TABLE VII

System	P.A.M.	P.L.M.	P.P.M.	F.M.	A.M.
P	1	$\frac{1}{2}\delta_L^2 B/p$	$\delta_P^2 B/p$	$3f_D^2/f_A^2$	1

In this table and below the meanings of the symbols are as follows—

p = pulse length, f_P = pulse repetition frequency, δ_P = peak deviation (in seconds) in P.P.M. and δ_L = peak deviation (in seconds) in P.L.M.

Both δ_P and δ_L will be taken to be approximately equal to p in Figure 81. As before, f_D is the frequency deviation in F.M. and f_A is the audio pass-band.

Second, the "threshold value" is required. This Jelonek gives as $12{\cdot}25\ Bpf_P/f_A$ for the case where the mean power is the same as in the A.M. system. Third, the slope of the curve for small signal/noise ratios must be known. For this Jelonek deduces a value of about 6 db/db, but points out that it will vary with the system parameters. For the purpose of presenting Figure 81, B has been taken as 2 Mc/s, p as 10 microseconds, f_P as 10 kc/s, and δ as $9\frac{1}{2}$ microseconds. The reader is referred to Jelonek's paper for a discussion of the effect of varying the parameters.

The improvement shown here for pulse systems is due to "slicing" the pulses before they enter the demodulator. This removes a considerable fraction of the associated random noise.

7.1.4. *General Modulation Comparisons.* A study of the modulation characteristics briefly described in the previous paragraphs leads to the conclusion that, judged on improved performance per unit input power, frequency modulation is always to be preferred. The response curve is, however, only one of many features of a communication system. An overall system comparison shows that to each modulation method there corresponds a particular domain in which it is technically superior.

Before leaving the subject, it should be noted that the signal/noise improvement for each method can be given a different meaning. The signal/noise (expressed as an amplitude ratio) may be said to measure the intelligence which can be carried by the radio link. This is strongly suggested by the thought that a voltage signal/noise ratio of n means that the system can distinguish between an amplitude $(n - 1)$ times the R.M.S. noise value and an amplitude n times the noise value. Virtually it enables the system to count with an accuracy of 1 in n. This is in very many cases a measure of the rate at which intelligence can be transmitted. Let it be assumed, therefore, that a given rate of dealing with intelligence can be measured by a given signal/noise in volts. Consider now an F.M. system. It has been shown that a signal/noise improvement of $\sqrt{3}f_D/f_A$ is obtained over a corresponding A.M. system at the expense of using a band-width $\pm f_D$ instead of $\pm f_A$. To bring the improved system back to the same signal/noise ratio as in the A.M. case the power must be divided by $3f_D^2/f_A^2$. This gives the following table which has been extended to take account of pulse systems also. In considering pulses it has been assumed that $f_P = 2f_A$, that the mean pulse length $p = 1/2f_P$, and that $\delta \doteqdot 9p/10$. The bandwidth B is then

TABLE VIII

System	Band-width	Power
A.M.	$\pm f_A$	P_0
F.M.	$\pm f_D$	$P_0 f_A^2/3f_D^2$
P.L.M.	$\pm B = \pm 10/p$	$4P_0 f_A/B$
P.P.M.	$\pm B = \pm 10/p$	$2P_0 f_A/B$

of the order $10/p$ in each case. The table shows the power and band-width necessary to produce the same rate of transmission of intelligence as in a standard A.M. case. It is assumed throughout that the signal/noise ratio even with the diminished power is large compared with unity. Clearly in A.M., P.L.M. and P.P.M. (band-width $\times$ power) is a constant for a constant rate of transmitting intelligence. In the F.M. case (band-width)^2 $\times$ power is a constant. The difference is due to the nature of the demodulator. In pulse communication the latter is an integrator; in F.M. it is a linear transformer.

The general law which emerges from such considerations is that the rate of dealing with intelligence is a function not only of the band-width employed (as suggested by Hartley) but also of the power. In fact power and band-width are related so that if $f(P)$ is a function of the power and $\phi(B)$ is a function of the band-width, the rate of transmission of intelligence is proportional to the product $f(P)\phi(B)$. If I represents intelligence, then for F.M.,

$$\frac{dI}{dt} = k_F B \sqrt{P}$$

and for pulse modulation

$$\frac{dI}{dt} = k_P \sqrt{BP}$$

7.1.5. *Pulse Code Modulation.* Although pulse code modulation is not in the same category as the methods heretofore discussed, this is an appropriate point at which to introduce the characteristics of the newest communication technique. From the considerations of the last paragraph it will be realized that it is possible in principle to reorganize the data of the intelligence to be transmitted in many ways so as to use a greater or smaller band-width. The essence of P.C.M. is first to translate the intelligence into a sequence of numbers and then to codify each number in such a way that it may be represented by a series of n ON–OFF signals (pulses). By this method it is possible to transmit intelligence with high precision, using a small signal/noise ratio (and therefore a small power), but a wide band-width. The burden of obtaining precision is removed from the radio link and is placed upon the coding system.

P.C.M. can be understood most clearly from an examination of speech coding. It has already been pointed out (Section 5.4.1) that speech can be transformed into a sequence of pulses of varying amplitude by a process of sampling, and that it is quite satisfactory to sample the speech at a recurrence frequency slightly greater than $2f_A$. In practice 8 kc/s is often used. We can then regard the speech as being translated into a series of "amplitudes" each of which could be represented by a number, and therefore the first part of the procedure in P.C.M. has been accomplished. Let us suppose now that the accuracy with which we wish to read off these numbers is to one-hundredth of the maximum amplitude. To do this the adventitious variations in amplitude due to noise must be at most one hundredth of the maximum amplitude also, and this corresponds to a signal/noise ratio of 40 db. One way of sending the speech with the requisite degree of accuracy would be to employ a band-width of 8 kc/s and a transmitter of such power that a 40 db signal/noise ratio would be assured.

The alternative presented by P.C.M. is to count the number of units in each pulse, taking the maximum as 100, and to transform each number into an n-pulse code, employing usually the binary system of numbers. Then each pulse, when coded, has become a sequence of ON–OFF pulses, and it is only necessary for exact reception that each ON pulse should be large compared with the noise. Hence the 40 db signal/noise at the output can be obtained by using a signal/noise ratio of (let us say) 10 db in the radio (or line) link. Clearly, this procedure economizes in transmitting power, but, since each of the original "sampling" pulses must be sent as n ON–OFF pulses, either the speed of sending must be divided by a factor n, or, alternatively, the band-width must be multiplied by a factor n. For example, if a 5-symbol code is used, numbers up to $2^5 = 32$ can be sent, giving a signal/noise of approximately 30 db. This is achieved at the expense of a band-width of 5×8000 c/s $= 40$ kc/s, using a signal/noise of 10 db in the link.

In the general case, if the band-width is $n \,.\, 8000$ c/s and it is doubled, the power can be diminished to $1/2^{2n}$ of its value, retaining the original signal/noise ratio. For systems requiring a very high signal/noise, this method of communication is much more economical in power and band-width than any other.

7.2. **Propagation.**[5] One of the factors determining overall system performance is the mode of propagation of the U.H.F. energy in the vicinity of the earth's surface. It is not proposed in this section to give a detailed account of the quantitative methods whereby it is possible to estimate the signal attenuation due to the medium. Rather some of the more important effects will be mentioned with the object of exhibiting their influence on equipment design.

7.2.1. *Frequency Factor in Free Space.* One of the great advantages of working in the U.H.F. region is that a very efficient collector or emitter of radiation can be made in a reasonably small space. It must not be forgotten, however, that the wave-length factor can operate in another way, detrimental to the higher frequencies. If it is necessary to receive radiation equally from all directions in the horizontal plane, then without employing a complex mechanical structure the resonant dipole is the most efficient aerial. If two transmitters of frequencies f_1 and f_2 and of equal power are situated at equal distances in free space from two receiving dipoles resonant to the same two frequencies, then the power collected by each is inversely proportional to the square of its frequency. This scale factor, due to the geometry, operates to the detriment of the higher frequency, so that, other things being equal, a link requires more power for a given signal/noise ratio as the frequency is increased. The effect can only be of importance in mobile communications, and even then may be wiped out by some other frequency-sensitive factor.

7.2.2. *Reflections from the Earth's Surface.* In many U.H.F. systems reflections from the earth's surface play a part. The two waves, direct and reflected, add vectorially at the receiving aerial, which must be regarded as being in a spatial interference field. The pattern of the latter is not of much interest in communications unless the amplitudes of the two waves are of the same order of magnitude. This usually occurs only when the receiver is near the horizon of the transmitter. The effect has been examined in detail at a wave-length of 6 mm.[3] and the pattern was found to agree well with theory. In air-ground communication in the 120 Mc/s band the effect has also been found. It is particularly important when the aircraft is at a great altitude.

7.2.3. *Curvature of the Earth.* The dominant propagation effect at U.H.F. is due to the curvature of the earth. As in the case of light, the receiver, moved too far from the transmitter, falls into the shadow of the earth where the diffracted ray is rapidly attenuated. This effect is so marked that, to a first order of precision, the range of a transmitter is the distance to the horizon, so long as the power radiated is greater than a certain minimum. This effect is virtually independent of frequency in the U.H.F. region.

Consequently, it is of considerable importance to obtain the maximum elevation above the earth's surface for both transmitting and receiving aerials when the maximum range is required. Usually the decrease of power due to lengthening the feeder is smaller than the increase in signal strength due to increasing the aerial height.

7.2.4. *Atmospheric Conditions.* This is second in importance only to the curvature of the earth. The normal atmosphere is one in which the dielectric constant diminishes uniformly with increase of height above the earth's surface, but the change in dielectric constant per hundred feet is very small. Due to the presence of water-vapour in the atmosphere and to local variations of temperature and pressure, the normal atmosphere, like the average man, is a scientific standard which has no real counterpart. Even in temperate latitudes enormous fluctuations in the electromagnetic properties of the lower atmosphere are possible, so that radio waves may, for example, be trapped between the earth and a layer of air of abnormally low dielectric constant, and may proceed as in a wave-guide far beyond their "normal" range. This so-called "duct" effect is even more common in the tropics, where U.H.F. ranges equal to three times the optical range are not unusual. The reverse effect in which radiation is bent away from the earth occurs occasionally in Arctic regions. This effect leads to reduced ranges in the vicinity of the earth's surface and may easily be the limiting factor with long-distance beamed links in which the signal/noise ratio must not fall below a certain high level. All these phenomena are liable to fairly rapid change, giving the same effects as normal fading at lower frequencies, but in a well-designed system there is no need to fear their effects.

7.2.5. *Obstructions.* It is *a priori* obvious that U.H.F. waves will be either reflected or absorbed by most solid objects and that a sharply-defined shadow will be cast by objects the linear dimensions of which are large compared with a wave-length. Consequently, in any U.H.F. link it is necessary to ensure an unobstructed atmospheric path between transmitter and receiver, if the optimum performance is to be obtained. Occasionally, use can be made of reflections to surmount or to evade obstacles, but such cases are a small minority. Where all-round communication is required, as in mobile working, it is necessary to regard all objects in the vicinity of the aerials as obstacles. These can produce well-marked indentations in the polar diagram, the effects being narrower and deeper at the higher frequencies.

7.3. **Aerials.**[6] Aerials for use at U.H.F. may be classified in many ways. There is so much diversity among them, and each aerial problem is so unique, that a general statement of what can be done is all that will be attempted here.

7.3.1. *The Half-wave Dipole.* In addition to having the status of a standard against which other aerials may be measured, the resonant dipole is by far the simplest and most popular U.H.F. aerial for all-round propagation or reception. The polar diagram in the plane of the aerial is such that an adequate fraction of the power is radiated over an angle of $\pm$ 30° about the plane of symmetry, and the radiation in the direction of the aerial itself is very small indeed.

The problems associated with a dipole are twofold. First, it is necessary to match it to the source of power (in the case of a trans-

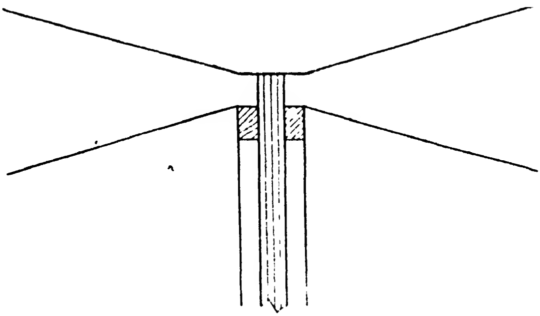


Fig. 82. Axial section of a biconical horn radiator

mitter) or to the sink of power (in the case of a receiver). Second, it is necessary to isolate the dipole from its cable in such a way that the cable will not itself radiate, for the polar diagram of the cable + aerial will be that of an aerial of many wave-lengths if this precaution is not taken. These two considerations lead to certain mechanical complications, but an elegant answer is always possible, unless the frequency band to be covered by the aerial is too wide.

7.3.2. *The Biconical Horn.* In many cases a sharp beam is required in the vertical plane; it is necessary to concentrate the energy flow into a narrow sheet, the medial plane of which is coincident with the plane of symmetry of the aerial. This calls for some modification of the dipole structure, and the most commonly used modification is the biconical horn. This is illustrated in principle by Figure 82 in which no attempt has been made to keep elements to scale. The angle of the beam, defining its width, is largely determined by the lengths of the cones, measured in wave-lengths, and the structure becomes unwieldy for gains greater than two at wave-lengths greater than 20 cm.

7.3.3. *Reflectors.* Placing a reflector behind a dipole directs all the energy to the front and by suitably shaping and positioning the reflector, exceedingly narrow conical beams may be obtained. In radar and communication systems in the centimetre wave-band, aerial gains of 100 to 300 are common, corresponding to beams of half-angle 8 to 3 degrees. The reflectors used are of circular or parabolic cross-section. This does not represent the limit of what can be done by any means. Using polished reflectors of sufficiently large diameter, accurately positioned relative to their collecting or energizing elements, gains of the order of 1000 are feasible in the same region of the spectrum. The formulae for aerial gain and for beam half-angle are as follows—

Let G = aerial gain relative to a uniform spherical radiator.
Let A = aperture area of the (paraboloidal) reflector.
Let λ = wave-length.

Then, if the aerial is adjusted for maximum gain,

$$G = \frac{4\pi A}{\lambda^2}$$

If G′ = aerial gain relative to a half-wave dipole,

$$G' \doteq \frac{3}{5}\frac{4\pi A}{\lambda^2}$$

and if the half-power point is reached by moving away from the axis of the reflector by angle θ, then

$$\theta = \sin^{-1}\left(\frac{0{\cdot}26\lambda}{a}\right)$$

where a = radius of the aperture, or $A = \pi a^2$.

7.3.4. *Dielectric Lenses.* A combination of an electromagnetic horn and a dielectric lens makes an excellent directive aerial of simple mechanical construction. The horn, consisting of a flared wave-guide, forms a rough beam and produces the appropriate impedance. The dielectric lens, placed at the mouth of the horn, defines the beam more exactly and prevents reflections. Such aerials, due to their simple mechanical construction, are very free from side and back lobes in their polar diagrams. These lobes are very troublesome when ordinary reflectors are used.

Much remains to be done to obtain the optimum aerial design from the mechanical standpoint. In America much attention has been given to dielectric lenses for use in high-gain, high-definition aerials. In this case the dielectric is employed in the form of a large number of strips which are prefabricated to an optical design. There is no reason why the efficiency of the lenses used for the propagation of light should not be improved upon, using this method.

7.4 **Radio Networks.** The use of the U.H.F. region of the spectrum for communication opens up a new field of possibilities. The practical exploitation of these possibilities will come slowly. It is the purpose of this section to describe some of the things which can be done and to draw attention to the advantages to be gained.

7.4.1. *Cross-band Telephony.* By this is meant the use by one station of a single U.H.F. frequency for transmission and reception, so that two stations have different U.H.F. frequencies but a common intermediate frequency. An example is given in Figure 66 of Chapter VI. Where the definition of the carrier or local oscillator frequency is done by means of a crystal and a series of multipliers, cross-band working is an attractive system. The saving in equipment and power due to the use of a single radio-frequency chain may be considerable, while the fact that it is possible to be lavish with local oscillator power leads to the maximum receiver sensitivity. Monitoring of the transmitter is simplified, and side-tone is obtained directly (when the distant carrier is present) from the modulation of the local oscillator. For a single U.H.F. telephony link cross-band working is very suitable.

7.4.2. *Variable Intermediate Frequency Reception.* An extension of the cross-band principle makes an R/T network at U.H.F. simple to set up and to operate. The principle can be exemplified by Table IX, where f_T is the transmitting frequency of each of the twenty stations in the net. It will be observed that all the transmitting frequencies lie between 1000 and 1010 Mc/s inclusive. If, therefore, the receiver local oscillator is at 990 Mc/s, all the stations lie in the 10 to 20 Mc/s band after one frequency conversion. Hence a good superheterodyne H.F. receiver covering this band can be fed from the converter and the resulting composite receiver, although it has no U.H.F. selectivity, will have the necessary adjacent channel selectivity to operate in the network.

TABLE IX

Station No.	1	2	3	4	5
f_T (Mc/s)	1000·5	1001·0	1001·5	1002·0	1002·5
f_A (c/s)	300	400	500	600	700
Station No.	6	7	8	9	10
f_T (Mc/s)	1003·0	1003·5	1004·0	1004·5	1005·0
f_A (c/s)	800	900	1000	1100	1200
Station No.	11	12	13	14	15
f_T (Mc/s	1005·5	1006·0	1006·5	1007·0	1007·5
f_A (c/s)	1300	1400	1500	1600	1700
Station No.	16	17	18	19	20
f_T (Mc/s)	1008·0	1008·5	1009·0	1009·5	1010·0
f_A (c/s)	1800	1900	2000	2100	2200

Each receiver may be fitted with a simple uniselector so that any of the twenty frequencies may be chosen at will.

To operate the network, let the exchange consist of a transmitter on 1000 Mc/s to which every subscriber's receiver is automatically tuned when the hand-set is in its cradle. Imagine also that the ringing system consists of a series of tones, one to each subscriber, as listed in Table IX under f_A. The subscriber's bell rings only if the receiver is at the correct frequency. Then if No. 4 wishes to call No. 19, he lifts his handset and waits. Exchange, connected *via* a switchboard to twenty H.F. receivers on the appropriate frequencies, answers on 1000 Mc/s, having had a visual indication of the presence of No. 4's carrier. No. 4 asks for No. 19. Exchange calls No. 19, using the appropriate 2100 c/s tone, and when No. 19 answers him, says "No. 4 wishes to speak to No. 19." Both have heard him and both then dial each other's number on their uniselectors. Thereafter, they talk to one another as on a telephone. Exchange monitors the first part of the conversation for safety and then withdraws his plugs. He knows when both go off the air by means of his carrier indicator lamp. Should another subscriber wish to make a call while the operator is busy, he will immediately hear the operator's speech and will wait his turn. In any case he cannot disturb the call which is being made, since his frequency is different. By ensuring that a subscriber's carrier automatically comes on when he lifts his handset, secrecy is achieved, since exchange will then ask "Number please!"

There are many variations of the system described above, the common technique being the use of a variable frequency amplifier after the frequency converter. It should be noted that in such systems almost all the frequency stability must be obtained by the fixed local oscillator. Variations in the frequency of the variable oscillator are relatively unimportant.

7.4.3. *Common Wave Duplex.* The use of the U.H.F. region gives the possibility of duplex working on a single frequency, although the technique requires development to be of universal application.

One method depends upon the application of the directional properties of the wave-guide and is consequently only practicable at frequencies at which the guide is of reasonable dimensions. The outstanding example is the so-called "magic tee" which is shown diagrammatically in Figure 83.

The cross-section of the guide must be such that the only important mode of propagation at the required frequency is $TE_{1,0}$. Then power injected into arm A can flow into C or D, but, so long as there are no reflected waves in either, power cannot find its way into B. Similarly, power injected into B cannot flow directly into A. Physically, the isolation of A from B is due to the definition of the plane of polarization: the rigid proof of the isolation is a matter of some difficulty, since it is not complete in any real system.

In practice, if the "tee" is properly matched, the isolation amounts to an attenuation of the order of 60 db between A and B at a spot frequency. Two inductive irises are usually placed in A and B to match the guides to the junction; alternatively, one of the irises may be replaced by a stub placed in the plane of symmetry of the device. It has been said that there must be no reflections in C and D. In particular, the load across each much be matched to the guide as closely as possible.

Now imagine the "tee" connected so that A goes to the transmitter, B to the receiver, C to the aerial and D to a dummy load.

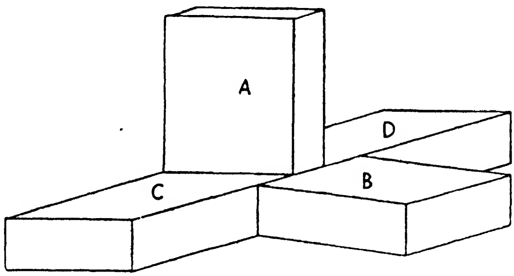


Fig. 83. A "magic tee," comprising a double "tee" junction in a piece of rectangular wave-guide

The tee can then be used to achieve single frequency duplex if a loss of 6 db in the system and an isolation of the order of 60 db can be tolerated. The loss of 6 db is due to the fact that half of the power must be dissipated in the dummy load both in transmission and in reception.

Should the isolation of 60 db be too small, an additional 60 db can be found by separating the transmitter and receiver in frequency. Usually, the adjacent channel selectivity of the receiver is great enough to obtain this additional 60 db rejection without materially reducing the rejection due to the "magic tee." The system is not then single frequency duplex, but it uses the minimum frequency separation and this is important in certain applications. It is probable that a "tee" isolation greater than 60 db will be achieved in the future when development of the device for communication purposes has been completed.

Another promising method of duplexing which is similar in principle is shown in Figure 84. This represents a four-terminal network, consisting of sections of concentric line, each of which is effectively $\lambda/4$ in length, and in which the impedances of opposite sections are equal. The ratio of the impedances of two adjacent sides is $\sqrt{2}$. An analysis of the network shows that it is balanced when matched loads are placed across C and D. A and B are then isolated from each other. This is a modification of the so-called "ring duplexer" which has its counterpart at low-frequencies. Again the attenuation

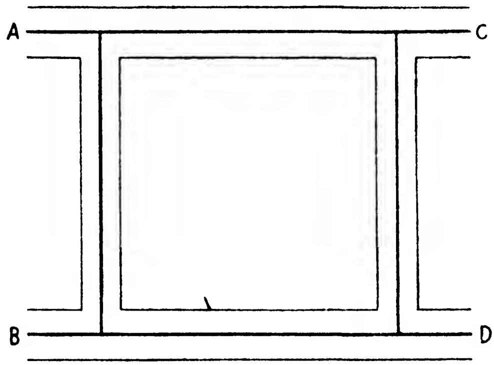


Fig. 84. Axial section of a concentric line duplex balancer

between A and B is not enough for genuine duplex, but a slight change of frequency is enough to produce a rejection factor of 120 db in a selective receiver.

Another simple method of achieving duplex operation with one

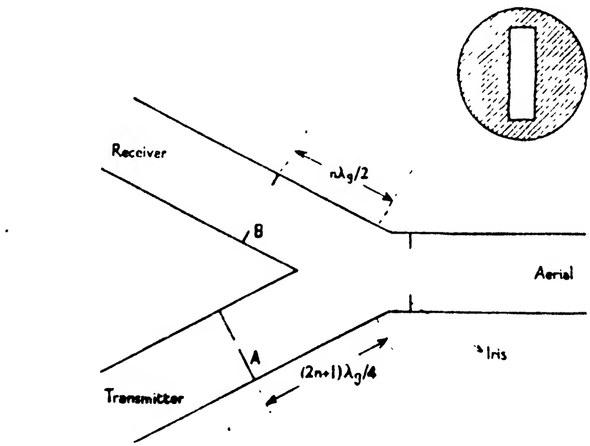


Fig. 85. Wave-guide duplexer in which the direction of propagation at the junction is determined by the plane of polarization, the transmitted and received waves having their electric field vectors at right angles

aerial at centrimetric wave-lengths is to use planes of polarization at right angles to each other for transmit and receive. The method can be explained in terms of Figure 85.

The plane of polarization of the transmitted wave is caused to be in the plane of the paper by means of the filter at A. A plan of the filter is shown in the inset to the figure. The direction of the electric vector for the $TE_{0,1}$ wave is parallel to the shorter side of the rectangular aperture. The wave-guide, which is of circular cross-section, is chosen so that it will only transmit the $TE_{0,1}$ mode. In the same way the filter at B only allows a wave to pass the plane of polarization of which is perpendicular to the paper. To avoid reflections, the filters must be placed at the appropriate distances from the junction. In the case of a wave approaching the latter from the aerial with its plane of polarization in the paper, the two impedances are in series, and to avoid reflection the impedance of the receiver arm should be zero. The filter B must therefore be $n\lambda_g/2$ from the "plane" of the junction. Reversing the direction of flow of the energy, all the power from the transmitter will then find its way into the aerial, if the conditions are as just stated. Energy fed from the aerial with the plane of polarization perpendicular to the paper sees the junction as a pair of parallel impedances. The impedance of the transmitter limb must therefore be infinite and the filter A must be placed at a distance $(2n + 1)\lambda_g/4$ from the plane of the junction. This ensures that the energy flows into the receiver.

In practice the duplexer just described has to be set up by trial and error, and imperfect matching of the aerial must be countered by a reactive diaphragm in the aerial limb. It has been reported that an attenuation between transmitter and receiver of 80 db can be achieved by this means. The duplexer is described in connection with a practical communication link by Lamont, Robertshaw and Hammerton.[4]

Finally, under the heading of common wave duplex, but applicable to all carrier frequencies at which voice communication can reach a good standard of performance, there must be mentioned the system known as Electronic Voice Operated Switching (E.V.O.S.). This can be briefly described as a device whereby the action of the voice on the microphone opens the local transmitter and shuts down the local receiver. Simultaneously, the radiated modulated carrier shuts down the distant transmitter and opens the distant receiver. By the incorporation of circuits to bring the modulation to a desired uniform level which is independent of the strength of the speaker's voice (V.O.G.A.D.) and by using the correct time constants throughout, E.V.O.S. can be applied to almost any transmitter and receiver. The latter must have a silencing circuit. When properly adjusted, E.V.O.S. is very effective. The first part of the speaker's first syllable must, of course, be chopped off, but the result is not significant. The few circuit clicks which accompany the automatic switching can be reduced to a level at which they are unnoticed, and the resulting system is not distinguishable from normal two-frequency duplex. It is, indeed, superior to the latter, in that a number of speakers can take part in a conversation simultaneously, so long as no one speaker is too selfish about his time "on the air." It is not true "duplex" in that only one voice channel is in operation at any instant.

7.4.4. *Broadcasting*. Broadcasting systems operating at micro-wave-lengths have many intriguing possibilities, some of which have already been exploited in U.S.A. It will, however, be some time before it has been finally determined which of the many alternative services is the most attractive. A fairly obvious application of U.H.F. to broadcasting lies in the colour television programme. Another simple case is the multi-programme broadcast, corresponding to multiplex communication, and providing a wide selection of programmes on a single carrier. Both of these applications would make use of the increased band available for modulation at these high carrier frequencies. But on the propagation side there are also remarkable possibilities. The exceedingly sharp cutoff at the optical horizon experienced by these micro-waves, coupled with the use of wide-band modulation, offers the possibility of a very sharply defined service area outside of which interference would be minimized. The exact definition of the service area could be improved by the use by the "listener" of a directional aerial and an insensitive simple receiver.

Imagine, for example, a broadcast service covering a circular area of radius 15 miles to aerials within sight of the transmitter. Suppose that the service consisted of six alternative programmes, radiated on a carrier at 2000 Mc/s and occupying a band of 2·5 Mc/s, using F.M. with a deviation ratio of 10. Let the radiated power be 1 kilowatt. Then, if the receiving aerial were a tuned dipole at the extremity of range (15 miles), neglecting reflected waves, the power delivered to the receiver would be

$$P = \frac{W}{4\pi r^2} \cdot \frac{5}{3} \cdot \frac{\lambda^2}{8} \quad . \quad . \quad . \quad . \quad (7.1)$$

where λ = wave-length = 15,

W = power of transmitter = 1000,

r = distance of receiver = 15 . 1760 . 36 . 2·54.

This gives $P \fallingdotseq 6 \cdot 10^{-10}$ watt. But $kTB \fallingdotseq 1{\cdot}3 \cdot 10^{-14}$, taking $k = 1{\cdot}4 \cdot 10^{-23}$, $T = 300$ and $B = 3 \cdot 10^6$. A noise factor of 20 db would be very conservative at 2000 Mc/s, giving a signal/noise ratio at the input of 25 db. An aerial gain of 15 db would be easy to achieve, raising the signal/noise ratio to 40 db, while the effect of the F.M. noise triangulation would be to improve this ratio at the output to a figure for each channel of 60 db. The radio-frequency units of the receiver need only consist of a tuned circuit (resonator) and a silicon crystal rectifier. The video equipment would necessarily be complex to accommodate the six alternative programmes.

It should be noted that when wide-band modulation is used, as in this case, the frequency stability required is not of a very high order. If the transmitter is adequately defined, the receiver need only be stable to $\pm$ 1 part in 4000. Although this would be difficult to achieve in a fixed circuit, a resonator tuneable over a small range would cope with the situation quite adequately, assuming that it were retuned from time to time as in the case of the more orthodox broadcast receiver.

7.4.5. *Point-to-Point Communication: Relays.* It is in the construction of radio links between fixed terminals that the ultra high frequencies will find their greatest use in the future. For this purpose the micro-wave portion of the spectrum is supreme, giving all the advantages associated with broad-banding with the certainty of using aerials of moderate size with gains which can be measured in thousands.

Taking once more a simple example, imagine that it is required to construct a 24-channel radio-telephone link over an optical distance of 60 miles. This might be required to bridge a narrow sea, or to cross difficult country. Indeed, as the techniques develop, such links may replace all cable paths, particularly in those countries where private enterprise provides its own communications. In this example, using a frequency of 3000 Mc/s, and assuming that a signal/noise ratio of 50 db is required, it is possible to determine the aerial gain which would be used and the transmitter power which would be necessary. Twenty-four speech channels could be accommodated in a band $\pm$ 150 kc/s wide, using frequency division multiplex. This band could be transmitted with a deviation ratio of 10 employing a band-width of $\pm$ 1·5 Mc/s. If transmitter and receiver were stabilized to one part in 10^4, a band-width of $\pm$ 2 Mc/s would suffice at the receiver. At 3000 Mc/s a noise factor of 10 db can be achieved, and a factor of 15 db is conservative for commercial operation. Therefore the noise level at the input, using the latter, would be $5 \,.\, 10^{-13}$ watts. The F.M. gain, due to triangulation (per channel) would be 20 db. If, therefore, a 50 db signal/noise ratio is required (per channel), the input signal/noise ratio need only be 30 db, i.e. the signal required is $5 \,.\, 10^{-10}$ watts. Employing equation (7.1), this gives $W = 30$ kW. An aerial gain of 30 db can be obtained, using paraboloids 6 ft. in diameter. If two such aerials are employed for transmitter and receiver, the required transmitter power is 30 milliwatts. That this figure is of the correct order of magnitude has been amply demonstrated by Western Electric Company, U.S.A., in their centimetre wave links, the parameters of which closely resemble those in the example just quoted.

The use of relay stations in which the transmitted signal may be amplified some 80 to 100 db transforms a single link into a trans-continental system. It is usual in such repeater stations to amplify the signal without complete demodulation. In time division multiplex systems the pulses may be reconstituted at the relay with a certain improvement in the signal/noise ratio, and in F.M. frequency division multiplex systems frequency compression and expansion may be used to give the effect of negative feed-back. But the use of relay stations is largely dependent upon the availability of reliable equipment which can be left unattended for relatively long periods. In this connection the U.H.F. valves are the limiting factor. Up to the moment of writing, the life of a U.H.F. valve is not satisfactory for unattended operation. This cannot be regarded as serious, however, for, if the economic pressure is there, the valves will undoubtedly be improved. The next few years should see the development of electronic equipment for use at micro-wave-lengths the reliability of which is as good as that already obtainable at lower frequencies.

REFERENCES

(1) Crosby, *P.R.I.E.*, 24, p. 898 (1936).
(2) *J.I.E.E.*, Pt. IIIA, 13, p. 533 (1947).
(3) Lamont and Watson, *Nature*, 158, p. 943 (1946).
(4) *Wireless Engineer*, 24, p. 323 (1947).
(5) "Propagation of Short Radio Waves," *M.I.T. Radiation Laboratory Series*, McGraw-Hill (1947).
(6) "Microwave Antenna Theory and Design," *M.I.T. Radiation Laboratory Series*, McGraw-Hill (1947).

INDEX