FOUNDATIONS OF WIRELESS

A. L. M. SOWERBY
M.Sc.

THIRD EDITION

Revised by
M. G. SCROGGIE
B.Sc., A.M.I.E.E.

FOUNDATIONS OF WIRELESS

By

A. L. M. Sowerby, M.Sc.

THIRD EDITION

Completely revised by

M. G. Scroggie, B.Sc., A.M.I.E.E.

Reprint January 1943

WIRELESS WORLD

ILIFFE & SONS LTD

DORSET HOUSE, STAMFORD STREET, LONDON, S.E.1

BOOK PRODUCTION WAR ECONOMY STANDARD

THIS BOOK IS PRODUCED IN COMPLETE CONFORMITY WITH THE AUTHORIZED ECONOMY STANDARDS.

MADE AND PRINTED IN ENGLAND
AT
THE CHAPEL RIVER PRESS, ANDOVER, HANTS
Rp. 12/42

PREFACE

ON MATHEMATICS

THIS book is non-mathematical in treatment, but nevertheless algebraic formulæ are quite freely used. The justification for this is simply that in no other way can information be recorded quite so briefly, clearly, and simply. But it is realized that many readers may have had no occasion to use algebra since their school-days; the following few paragraphs, which attempt little more than an explanation of the meaning of letters used in place of ordinary numbers, may make it easier to recover the school-boy facility for understanding a formula.

An algebraic formula is an abbreviated instruction to perform an arithmetical process. How, for example, do we measure the speed of a car? If it goes 15 miles in half an hour, or 10 miles in 20 minutes, we spot at once that the speed is 30 miles per hour. How? By dividing distance gone in a given time by the time taken ($15 \div \frac{1}{2} = 30$; $10 \div \frac{1}{3} = 30$). When we recognize this, we have a means of showing anyone who does not know how to find the speed exactly what to do, irrespective of the actual values of the times and distances involved; we tell him to "Divide distance gone by time taken".

Probably we forget to tell him that if the answer is to be in miles per hour, distance must be measured in miles and time in hours—it seems too obvious. Yet, in seeing that 10 miles in 20 minutes equals 30 m.p.h., we have automatically regarded 20 minutes as one-third of an hour. (Dividing miles gone by time taken *in minutes* gives speed in miles per minute—one-half in this case.) Similar attention to the units of measurement, not necessarily automatic in all problems, is always needed.

If in our instructions we place the word "divide" by its mathematical symbol, our words take the form:

$$\text{Speed equals } \frac{\text{Distance Gone}}{\text{Time Taken}}$$

Here we have a brief convenient exposition of an extremely general type; note that it applies not only to cars, but to railway trains, snails, bullets, the stars in their courses, and anything else in heaven or earth that moves.

One small step further, and we are up to our eyes in algebra; let us write $S = D/T$, where S stands for speed, D for distance, and T for time taken.

Letters stand Proxy for Numbers

Observe that to say that S equals D/T is utterly meaningless unless we say what the letters are meant to stand for. Most people who fail to grasp the essential simplicity of algebraic expression do so because they think that the letters used have some meaning *in themselves*, and do not realize that they only stand for numbers as yet unspecified.

Those faced for the first time with an algebraic expression such as this often say, "But how *can* one divide D by T? Dividing one letter by another doesn't *mean* anything. If only they were numbers, now . . .". Well, of course, that is just what they are—ordinary numbers, only we don't yet know their exact values. But we do know that when these values are found, dividing one by the other will give us the answer we want. So, in place of leaving blanks for the figures ("Blank divided by blank" would be ambiguous, to say the least of it) we put in letters, carefully defined in meaning, to act as temporary substitutes. No question of "dividing one letter by another" ever arises; one waits for the numbers.

Instead of looking on "$S = D/T$" as an instruction for calculating the speed, we can regard it as a statement showing the relationship to one another of the three quantities, speed, time and distance. Such a statement, always involving the "equals" sign, is called by mathematicians an "equation". From this point of view S is no more important than T or D, and it becomes a mere accident that the equation is written in such a form as to give instructions for finding S rather than for finding either of the other two. Since they are equal, we can divide or multiply the two "sides" of the equation by any number we please without upsetting their equality; if we multiply both by T (using the ordinary rules of arithmetic, since S, T, and D are really unspecified numbers) we get "$S \times T = D$". Our equation now has the form of an instruction to multiply the time of the journey by the speed in order to find the distance gone. (Two hours at 20 m.p.h. takes us 40 miles.)

If we like to divide the new form of the equation by S, we get "$T = D/S$"—an instruction, now, to find the time consumed on a journey by dividing distance gone by speed. (Thirty miles at 20 m.p.h. would take $1\frac{1}{2}$ hours.)

It is important to note that these conversions of our original equation into new forms are independent of the meanings of the letters; the process is purely arithmetic, and consequently cannot give more information than the original equation contained. But such transformations are frequently made in order to twist the information provided into a form that will be more convenient when we come to put in the numbers for which the letters stand.

Other Symbols in Algebra

Wells's delightful episode of a tramp trying to read algebra—"Hex, little two up in the air, cross, and a fiddlededee"—reminds us that there are algebraic symbols other than letters. These, again, are only instructions to perform certain arithmetical operations on the numbers for which the letters stand.

For example:

ab	means	"Multiply a by b". Usually the tramp's "cross" is left out, and mere juxtaposition signifies multiplication. But the cross is restored and we write "$a \times b$" where its absence might produce ambiguity.

x^2 ("Hex, little two up in the air" — read by the initiated as "x squared")	means	Multiply x by x. x^n means n x's multiplied together.
$\sqrt{x}$	means	Take the square root of x, or find the number which, when multiplied by itself, makes x. Algebraically, we can say $(\sqrt{x})^2 = x$.
V_1, V_2, V_3		These "subscript" figures have no algebraic meaning; V_1, V_2, and V_3 are *single symbols*, each standing for a different V. They may be voltages at different parts of a circuit, numbered thus to distinguish them.

Unless the square-root sign ($\sqrt{\ }$) can be so called, we have not yet found a fiddlededee. Perhaps Greek letters belong to this mysterious class—several are in frequent use In particular "π" (read as "pi") is always used for the ratio of the circumference of a circle to its diameter; it is mentioned here because its meaning is almost always taken for granted. The corresponding numerical value is 3·1416 approximately (the decimal never ends), or about 22/7. Other "fiddlededees" will be defined, like English letters, when we come to them.

Practical Use of Symbols

Having defined our symbols, let us see how they work. A commonly-used wireless formula is "$\lambda = 1885\sqrt{LC}$", λ (lambda) being wavelength in metres, L inductance in microhenrys, and C capacitance in microfarads. The formula tells us that if we multiply (the numerical value in any particular case of) L by (the numerical value in that particular case of) C, take the square root of the result, and multiply that by 1885, we shall be rewarded by (the numerical value in that particular case of) the wavelength. Usually we say, more briefly, "multiply L by C", omitting the long-winded phrases in brackets. This, though really meaningless, is justified by the fact that we *can't* multiply L by C until we know what numerical values to take. Meanwhile we just write "LC" as an instruction to multiply as soon as that information is known. Note that the square-root sign is extended over both L and C; this means that the extraction of the square root is to be applied to the product, and not to L only.

Brackets have a similar effect in lumping together the letters or figures within them. To find, in any numerical case, the value of $a + b\left(\frac{m}{n} + p^2\right)^2$, we proceed thus: Square p, and divide m by n. Add the results, and square the sum so obtained. Multiply this by b, and then add a. Note that b, being outside the bracket, is not squared, but that it stands as multiplier to the whole term $\left(\frac{m}{n} + p^2\right)^2$. Note also that $-x \times -x$ (or $-(-x)$) equals $+ x$.

Examples of numerical substitution will be found in the body of the book, so none are given here.

Symbols for Verbal Convenience

It only remains to point out that when there is used a phrase like " the resistance R " it is not to be assumed that by virtue of some superior knowledge the writer is assured that this resistance *is* R, and that the reader has to accept that fact as one more of the unsolved mysteries of wireless. It only means that it is proposed to save space by using the symbol " R " to stand for " the numerical value of the resistance, whatever it may eventually turn out to be ", or perhaps for " the numerical value of the resistance, whatever may be the value we choose to make it ". Sometimes, indeed, the letter is just a handy label, meaning " the particular resistance marked R on the diagram ". Often it will combine these meanings, and E/R may stand for "E divided by the numerical value of the resistance marked R in Fig. So-and-so ".

Space is often saved also by using the " index notation " for very large or very small numbers. 10^6 means six tens multiplied together (see definition of x^2), which comes to one million. As an extension of this, 10^{-6} means " one divided by 10^6 ", or, in mathematical terms, " the reciprocal of" 10^6. It is, of course, one-millionth part. "$3{\cdot}2 \times 10^{-12}$" thus means "3·2 divided by one million million". The justification for this notation is that "0·0000000000032" is extremely difficult to read.

Note that $10^6 \times 10^6 = 10^{12}$, that $10^{12} \times 10^{-6} = 10^6$, and that $\frac{10^{12}}{10^6}$ is only another way of writing $10^{12} \times 10^{-6}$. Multiply, in short, by adding indices, and divide by subtracting them.

CONTENTS

CHAPTER

CONTENTS

PAGE

CHAPTER

CONTENTS

PAGE

FOUNDATIONS OF WIRELESS

CHAPTER I

AN OUTLINE OF BROADCASTING

1. What Wireless is Not

IN discussing the day's wireless programmes one might easily remark to a friend: "There's some good music on the air to-night". Perhaps the phrase is more American than English, but it will nevertheless serve as a text for discussion, because it suggests a point of view that must be utterly abandoned before even beginning to grasp the mechanism of wireless transmission.

"Music on the air" suggests that the transmitting station sends out music as a disturbance of the air, which is music as we understand it in every-day life. But a transmitter is not a super-megaphone bawling out music; its aerial emits no more sound than does an ordinary telephone wire. "Music" must therefore be sent out from a wireless station in some altered state, from which it can be converted back into ordinary audible music by the listener's receiving equipment.

Anyone who has watched a cricket match will recall that the smack of bat against ball is heard a moment after bat and ball are seen to meet; the sound of the impact has taken an appreciable time to travel from the pitch to the grandstand. If the pitch were 1,100 feet away from the observer the time delay would be one second. Yet it is found that a watch may be set with apparently perfect accuracy by a wireless time signal from New York, providing, of course, that we allow for the fact that Americans do not use Greenwich Mean Time. That time signal has hurtled across the Atlantic in about a fiftieth part of a second. Comparing this with the three hours that would be required by any air-borne impulse we are driven to the conclusion that wireless transmissions travel in some other medium.

In the light of these facts "music on the air" has resolved itself into a silent substitute for music, carried by something that is not air.

2. Nature of Wireless Signals

The clue to the real nature of wireless signals is given by their rate of travel, which is the same as that of light. Light is one of the many possible disturbances in a mysterious and rather debatable medium called the "ether of space"; besides light there exist both longer and shorter ether waves which do not affect the eye at all.

The shortest waves, a few millionths of an inch long, affect only the smallest things, and are used by physicists to evoke disturbances within the atoms of which matter is composed, or to peer into atomic structure. The longer waves, which may be many yards long, also act on objects of physical dimensions comparable with their own. In particular, they affect metallic objects, such as wireless aerials, for example, losing energy to them and setting up in them electric currents. All these waves, since they are all carried by the ether, travel at the same rate, which is about 186,000 miles per second.

3. Transmission and Reception

Natural processes are mostly reversible, so that the fact that ether waves of long wavelength set up electric currents in an aerial wire at once suggests that if by any means electric currents of a suitable kind can be made to flow in an aerial, that aerial will very probably radiate waves into the ether. In actual fact it does so, and recognition of this at once makes it evident that communication can be carried out between two points, even though separated by many miles, provided that we have some means of generating the currents at the transmitting end and recognizing them at the receiver.

The whole process is no more and no less wonderful than ordinary speech, during which air waves are set up by the motions of the speaker's vocal cords, transmitted over a distance of a yard or two by the intervening air, and reconverted into mechanical movements when they strike the listener's ear drum. The sequence " electric currents—electric waves—electric currents " is exactly analogous to the sequence " mechanical motions—air waves —mechanical motions ". Communication by air waves for which we use our own natural organs, seems merely commonplace ; communication by electric waves is still something of a novelty, because it is only in this century that man has learnt to build himself transmitting and receiving stations, which are the electrical equivalents of mouth and ears.

The long distances over which wireless communication is possible is a result of the natural properties of the longer ether-waves ; in communication by signal fires and heliograph the shorter (visual) waves have been used for generations for the sake of their ability to span greater distances than can conveniently be reached by waves in the air.

4. Waves

Of the various types of wave that we meet in daily life those formed when still water is disturbed are the nearest in character to the invisible air or ether waves. If we drop a stone into a pool and watch the resulting ripples carefully we shall observe that as they pass a twig or other small object floating on the surface they cause it to bob up and down. But the twig is not carried along bodily by the ripples.

The waves, therefore, do not consist of water flowing outwards from the point where the stone hit the surface, although they certainly give the impression that this is happening. As the twig shows, all that the water at any one point does is to move up and down rhythmically a few times before the wave dies away. The point is that nothing moves outwards from the centre but *energy* passed on from one part of the water to the next.

The behaviour of an air-wave is very similar. Suppose someone seated in the middle of a large room claps his hands. A listener seated against the wall will hear that hand-clap almost immediately. It is not to be imagined that the air suddenly compressed in the act of clapping has shot across the room to the listener's ear in that brief time. What has happened is that the body of air suddenly compressed by the clap has rebounded, compressing in the process the air immediately surrounding it. This, rebounding in its turn, has passed on the wave of compression in the same way until it has eventually reached the listener. All that has actually travelled across the room is *energy* in the form of compression of the air.

5. Frequency and Wavelength

In wireless work one is more largely concerned with rhythmic waves than with irregular disturbances like that caused by a hand-clap. A stretched string, which emits a definite musical note, gives rise to a more important type of air wave.

When such a string is plucked or bowed it vibrates in the manner indicated in Fig. 1*a*. The movement of the string is rhythmic in the sense that each complete *cycle* of movements, from the highest position of A to the lowest *and back again*, occupies the same period of time. Moreover, each of these cycles is exactly like the last in every respect save that as the vibration dies away the amplitude of movement of the string becomes progressively less.

Fig. 1*a* (left): A stretched string vibrates in a regular manner when plucked or bowed, giving rise to a musical note of definite pitch. The size of the weight W controls the tension of the string, and therefore the pitch of the note

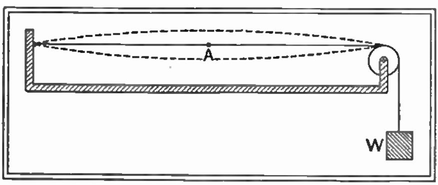
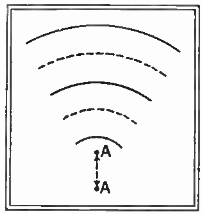


Fig. 1*b* (right): End view of the vibrating string at A in Fig. 1*a*. As it moves up and down over the distance AA it sends out alternate waves of compression (full line) and rarefaction (dotted line), which carry some of the energy of vibration to the listener's ear



The length of the time occupied by each cycle determines the pitch of the note heard ; if it is short, so that many vibrations take place each second, the note is high, while if it is long, so that only a few cycles of the movement occur in a second, the note is low. In scientific work of all kinds it is customary to specify a note in terms of the number of complete vibrations that occur in each second,

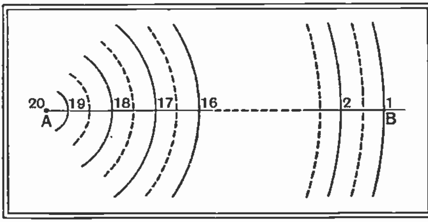
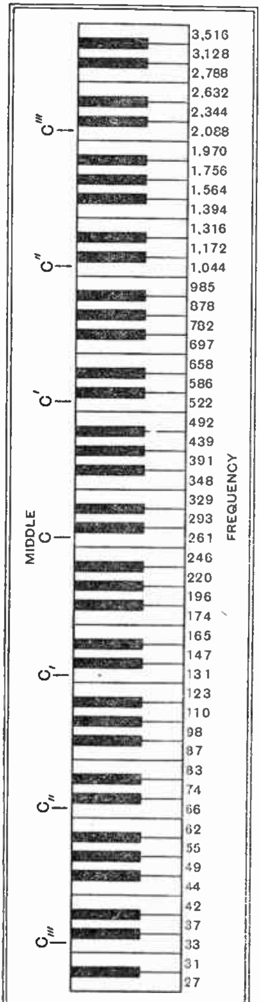


Fig. 2 : Twenty successive waves from the string. If the string is vibrating 20 times per second, the 1st wave has been travelling for one second by the time it reaches B, and the 20th is just leaving the string at A. Since sound travels 1,100 feet in one second, AB = 1,100 feet, and the distance between one wave and the next (wavelength) is 1/20th of 1,100 feet

this being known, for the sake of brevity, as the *frequency*.

Suppose the string vibrates at the rate of 550 cycles per second ; in each second it will send out 550 compressions and 550 rarefactions of the air. The rate at which the wave that these compose will travel forward depends only on the medium through which it is passing ; in air the velocity is about 1,100 feet per second. If we imagine that the string has been in vibration for exactly one second the wave corresponding to the first vibration will have reached a distance of 1,100 feet from the string just as the last wave (the 550th) is leaving it. There are, therefore, in existence 550 complete waves extending over a distance of 1,100 feet, from which it is very evident (compare Fig. 2) that each wave must be two feet long. If the string had executed 1,100 vibrations in the same period, the first would still have travelled 1,100 feet in the second of time occupied, and there would have been 1,100 complete waves in the series —each, therefore, one foot long. Since the velocity of sound in air is constant the higher frequencies correspond to the shorter wavelengths, and vice versa. It is specially



to be noticed that it is the *frequency* of the vibration that is fundamental, and that the wavelength is a purely secondary matter depending on the velocity with which the wave travels. That it really is frequency, and not wavelength, that settles the musical note heard can be shown by sending a sound through water, in which the velocity is 4,700 feet per second; the wavelength corresponding to a 550-cycle note is much greater than in air, but the pitch, as judged by the ear, remains the same as for a 550-cycle note in air.

The range of musical sound with which a wireless engineer has to deal runs from a low note of frequency about 25 cycles per second to a high note of frequency some 8,000 cycles per second, since if this range is fully reproduced music is sufficiently natural to give real pleasure to even a critical listener. The musical frequency-scale of Fig. 3 indicates, for reference, the frequencies corresponding to various notes.

Fig. 3: Showing the frequency corresponding to each musical note. Harmonics (multiples of the fundamental frequency shown) give notes their distinctive character; hence the need to reproduce frequencies outside the range of music as written

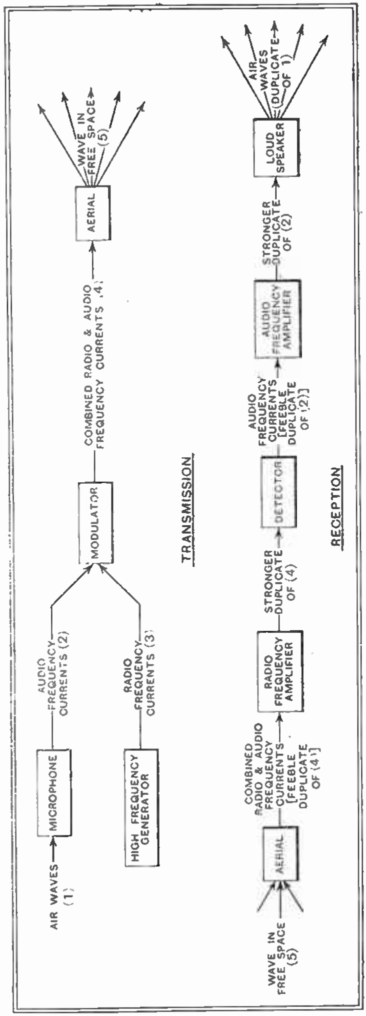


Fig. 4 : Schematic outline of broadcasting, showing how air-waves in the transmitting studio are duplicated, after many transformations, in the listener's home. Many stages of amplification have been omitted, for the sake of simplicity, from the diagram of transmission

6. Wireless Waves

When we turn to the production of the wireless waves, by whose aid music is transmitted from place to place, we find frequencies of a very different order. These waves, as has already been mentioned, are set up by the surging to and fro of electric current in the aerial of the transmitter. Since the flow of electric current does not involve the movement of material objects, as does the vibration of the strings and reeds used in music, there is no great barrier to the production of very high frequencies indeed. If the current in the aerial vibrates at such a rate as to complete the double motion a million times in a second, it is *oscillating* at quite an ordinary *radio-frequency*. In such a case the surging current sends out into the ether a wave which has the electrical equivalent of compressions and rarefactions, the "compressions" following each other every millionth of a second.

Being a wave in the ether, our wireless wave travels at the invariable speed of all ether waves, 300,000,000 metres, or 186,000 miles, in each second. If, during one second, one million complete waves are radiated by the aerial, then at the end of that time the first wave has travelled 300 million metres and the millionth is just leaving the aerial. Each wave, therefore, is 300 metres long. Just as in the case of sound, a lower frequency of electrical oscillation in the aerial will give rise to fewer waves each second, though the distance over which the waves emitted in one second will stretch remains the same. The waves, therefore, are longer. In symbols, the relationship is $\lambda = \frac{300{,}000{,}000}{f}$, where λ = wavelength in metres and f = frequency in cycles per second.

In dealing with sound, frequency is always used to specify the pitch of the note; in wireless matters both frequency and wavelength are in common use. Since in this book we shall be much less concerned with the waves themselves than with the rapidly oscillating electric currents from which they are born and to which they give rise, we shall exhibit a definite bias towards the use of frequency rather than wavelength, on the grounds that the specification of wavelength is really meaningless except when considering a wave in free space.

7. From Studio to Listener

With a knowledge of the nature and relative frequencies of sound and wireless waves we can trace through, in the broadest outline, the whole process of broadcast transmission and reception. It is summed up, with almost ludicrous absence of detail, in the crude scheme of Fig. 4.

We begin in the studio, where we will imagine that an orchestra is playing a symphony. The result, brutally ignoring the æsthetic side, is a complicated mixture of air waves. These impinge on the diaphragm of a *micro-*

phone, and this diaphragm, being thin, light and flexible, takes on exactly the movements of the air in which it stands. The task of the microphone is to convert these movements of its diaphragm into movements of electricity, just as though the wire leading from it were a pipe filled with water pushed to and fro by the diaphragm.

The complicated air-waves are thus eventually translated into exactly corresponding movements of electricity, so making a complex mixture of currents at frequencies which may lie anywhere within the range 25 to 8,000 or more cycles per second. They cannot be radiated from the aerial in their present form, partly because they are too weak and partly because the frequencies they represent are far too low to radiate well.

From another source a single regularly oscillating current, of a frequency suitable for wireless purposes (150,000 to 1,500,000 or more cycles per second) is produced, and the currents from the microphone are superposed on this in such a way that they mould (or *modulate*) it into their shape. The mixture is finally fed to the aerial, so that the wave sent out bears upon it, in the form of variations of strength, the impress of the currents derived from the microphone. These are then carried, in their new form, to any point on the globe to which the wireless wave itself can reach.

It has already been pointed out that a transmitter sends out a silent substitute for music; this complex wave is that substitute.*

On striking an aerial this wave is partially absorbed by it, the energy so abstracted from the wave serving to set up in it a current which is an exact replica in miniature of the far more powerful current surging back and forth in the aerial of the transmitter. If the received signals† are very feeble, as they may be if the transmitter is distant or the aerial small, the first need is to strengthen them without changing their character. This is done by a *radio*

* The nature of this wave is discussed in detail in Chapter 8.

† In the absence of a better word, the meaning of "signals" has been extended to include the electrical equivalents of speech and music.

frequency amplifier (Chapter 11.) When sufficiently amplified the signals are passed to a *detector* (Chapter 9), which sorts out from the complex current representing the wave as a whole those parts of it which are directly due to the original music, rejecting those more rapidly oscillating currents which, in enabling the music to be transported from transmitter to receiver on the wings of a wireless wave, have now done all that is required of them.

The currents we now have left are as exact a copy of those given by the microphone in the studio as can be had after so many transformations; they only require to be magnified up by another valve or two until they are strong enough to operate a loudspeaker (Chapter 13). To this they are accordingly passed, where they serve to push and pull a diaphragm (usually of paper) in such a way that its movements are a mechanical replica of the movements of the electric currents supplied to it. The diaphragm of the speaker thus performs the same movements as did that of the microphone a fraction of a second earlier; in doing so it sets up in the listener's home air waves which are, as nearly as may be, identical with those produced by the orchestra.

CHAPTER 2

ELEMENTARY ELECTRICAL NOTIONS

8. Electrons and the Electric Charge

THE exact nature of electricity is a mystery that may never be fully cleared up, but from what is known it is possible to form a sort of working model or picture which helps us to understand how it produces the results it does, and even to think out how to produce new results. One of the most startling conclusions is that all matter is electricity. The reason why the very existence of electricity went unnoticed until comparatively recent history, although it lay all around, is that it consists of two opposite kinds which normally exist in equal quantities and cancel one another out. For want of a better description, these kinds are called positive (+) and negative (−). In arithmetic $+1-1$ equals just nothing; and so although both positive and negative electricity produce very remarkable effects when they are separate, a combination of equal quantities is just ordinary matter—air, water, wood, and so on—without any apparent electrification.

One of the chief characteristics of electricity is a very strong attraction for the opposite kind. That is why they are so seldom found separated. Once each tiny particle of electricity finds its mate the two are so satisfied with one another that they exert no influence outside themselves.

It appears that the only difference between different kinds of stuff is the number and arrangement of the particles of electricity that form their atoms. The gas hydrogen, for example, is the simplest known substance; each atom of it consisting of one positive of electricity, called a *proton*, with one negative particle, an *electron*, revolving around it, like the moon round the earth, except that it does so at an incredible speed. Other atoms are of more complicated structure; for example, there might be four protons and two electrons at the centre, with two other electrons revolving round them (Fig. 5).

Normally the total numbers of protons and electrons are equal, but the one or more outer electrons can be detached by various means—heat, light, friction, etc.—leaving an atom deficient in negative electricity and as a result showing all the characteristics of positive electrification, or, as it is usually called, a positive charge. The detached electron, or any piece of matter harbouring it, displays every evidence of a negative charge, including an attraction for positive charges.

The space between them, across which this attraction is

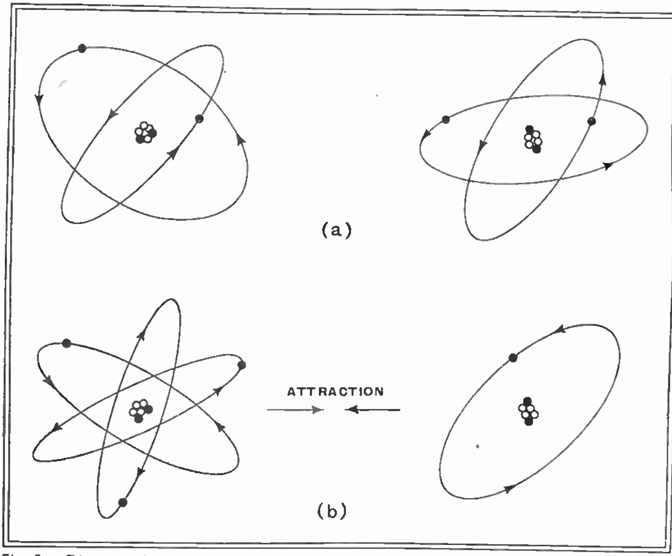


Fig. 5 : Diagram showing (a) imagined structure of two neutral atoms, each containing the same number of protons (white) and electrons (black), which cancel out so far as electrical effects are concerned. *(b)* shows the same two atoms after one has gained an electron at the expense of the other, and they now show electrical attraction. Two atoms, both with a surplus or a deficiency of electrons, repel one another

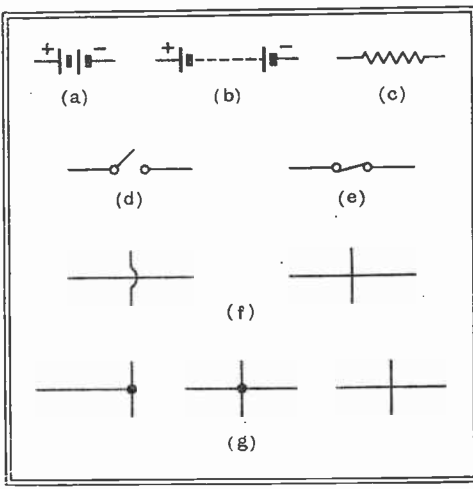
exerted, is said to be subject to an *electric field*. The greater the number of opposite charges, and the closer they are, the more intense the field and the greater the attractive force. The force tends to drive electrons, which are much more mobile than their positive counterparts, from places where they are in the majority to places where they are relatively fewer, until there are equal proportions on both sides, when the attractive force ceases. The action is rather like that of water, which always tends to flow between any two points at different levels until the levels are equivalent.

It needs a certain amount of energy to force electrons to go to a place where they are already in a majority, against the repulsion between like kinds; and this energy is stored ready to be released if the attraction for electricity of the opposite kind is allowed to assert itself, reuniting the electrons with their positive mates. Water power, too, can be stored, by pumping it up to a higher level against the force of gravity, and can then be used to drive a water-wheel or turbine in its descent.

9. Conductors and Insulators

For this flow to take place a pipe or other channel must be provided, while for an electron flow an electrically conducting path must be provided between the two points. In a conducting material electrons are very readily detached from their parent atoms, so that if a wire is stretched between two oppositely charged bodies, electrons can enter the wire at one end and cause a displacement of free electrons all down the wire, resulting in the emergence of an

Fig. 6: Some conventional signs used in constructing electrical diagrams. (*a*) A battery of few cells, used for filament accumulator or grid battery. Two cells are shown, making either a 4-volt accumulator or a 3-volt dry battery. (*b*) A battery of many cells, e.g., a high-tension battery. (*c*) A resistance. (*d*) A switch, shown open. (*e*) A fuse. (*f*) Wires crossing: the sign on the left is more usual. (*g*) Wires joining: the "dot" is generally used. Note that a simple line always indicates an electrical connection of negligible resistance



equal number of electrons at the other. Picture a long pipe of very wide bore, already filled with water. If an extra teaspoonful of water is forced into it at one end a teaspoonful will emerge at the other—but not the same actual water. If milk had been forced in instead of water, water would still have emerged. In the same way, the wire in its normal state must be pictured as already filled with electrons, all in continuous random movement from atom to atom. The passage of electricity through the wire amounts to no more than the superposition upon this vast random movement of a trifling drift in one direction ; the emerging electrons may only have moved a thousandth of an inch.

If the atoms of a substance have their electrons so firmly fixed that this exchange is not possible, the material will not conduct ; it is called an *insulator*. All metals are *conductors* ; to the class of insulators belong ebonite, bakelite, rubber, the silk or enamel covering on wire, and, indeed, most non-metallic substances.

The flow of electrons through a conductor constitutes a *current of electricity*.

10. Fundamental Electrical Units

So far we have considered the current as originating from a body which has a small and temporary excess of electrons ; when the charge is dissipated the current must inevitably stop. There are, however, certain appliances, such as batteries and dynamos, that have the power of continuously replenishing the surplus of electrons ; so that if the two terminals of a battery are joined by a conductor

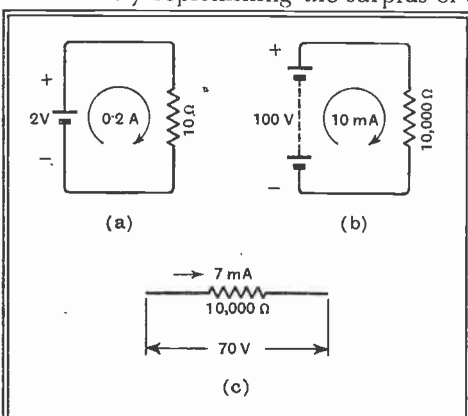


Fig. 7 : Circuits, illustrating Ohm's Law, constructed from symbols of Fig. 6. *a* and *b* show the application of Ohm's Law to a complete circuit, including the E.M.F. of the battery. (c) The application of the Law to part of a circuit ; if 7 mA flows through 10,000 ohms the voltage across the resistance must be 70 volts

a current flows continuously, or at least until the chemicals in the battery, which are responsible for the action, are used up. The difference in electrical level or pressure tending to drive a current through any continuous path, or *circuit*, leading from one terminal to the other, is called the electromotive force or E.M.F. and is measured in *volts*. The current that flows might very reasonably be measured in terms of the number of electrons passing from the battery into the circuit each second, but the electron is so extremely small that such a description of any useful current would lead to inconveniently large numbers. In consequence, it has become customary to take as the practical unit a body of about six million billion (6,000,000,000,000,000,000) electrons. This unit is called the *coulomb*, and is a unit of *quantity of electricity*, just as the gallon is a unit of quantity of water.

Just as one might speak of a flow of water of so many gallons per second, one can quite correctly describe an electric current as so many coulombs per second. Such a description, however, is rather cumbersome for frequent use, and the composite unit coulombs-per-second, as a measure of the rate of flow of electricity, is replaced by the more briefly named unit, the *ampere*. The statement that a current of one ampere is flowing means that one coulomb of electricity, or about 6×10^{18} electrons, flows past any point in the path of the current in each second.

Water, driven through a pipe by a constant pressure, will flow at a rate depending on the frictional resistance between the water and the inside of the pipe. Further, a pipe of large diameter will offer less resistance than one of small bore, and so will carry a larger flow at any given pressure. In just the same way, the magnitude of the current of electricity driven through a conductor by a battery depends on the electrical resistance offered by that conductor to its flow, and a thick wire offers less resistance than a thin one of equal length.

Circuits, especially the more complex ones, are more easily grasped from a diagram than from a description in words. Fig. 6 shows some of the conventional symbols from which electrical diagrams are constructed. A circuit element designed to provide resistance is known as a *resistor*. Each type of component—battery, resistance, switch, etc. —has its own sign, and the way they are joined up to make the complete circuit is indicated by heavy lines representing the wiring. A wire is always supposed to provide an electrical connection of negligible resistance. Other symbols will be introduced into diagrams as they are needed.

11. Ohm's Law

Electrical resistance is measured in units called *ohms*.

The relationship between E.M.F., resistance, and current is the most fundamental and important quantitative relationship in electrical science; it is known, in honour of its discoverer, as *Ohm's Law*.

Ohm's Law may be written as:

$$\text{Current in amperes} = \frac{\text{E.M.F. in volts}}{\text{Resistance in ohms}}$$

or, using the usual single-letter abbreviations for the three quantities, as $I = E/R$.

It will at once be seen that if for any particular case any two of these quantities, voltage, resistance, and current are known, the third can immediately be found. If, for example, we have a 2-volt accumulator connected to a length of wire having a resistance of 10 ohms (Fig. 7), the current flowing will be 2/10ths of an ampere. If the resistance had been only half this value, the current would have been twice as great, and it would have had this same doubled value if the original resistance had been retained and a second accumulator cell had been added to the first to make a total E.M.F. of 4 volts.

Taking another case, we might find, in investigating the value of an unknown resistance, that when it was connected across the terminals of a 100-volt high tension battery a current of 0·01 ampere was driven through it. Twisting Ohm's Law round into the form $R = E/I$, we get for the value of the resistance $100/0{\cdot}01 = 10{,}000$ ohms. Alternatively, we might know the value of the resistance and find that an old battery, nominally of 120 volts, could only drive a current of 0·007 ampere through it. We could deduce, since $E = I \times R$, that the voltage of the battery had fallen to $10{,}000 \times 0{\cdot}007 = 70$ volts

12. Practical Units

No wireless engineer would ever describe a current as 0·007 ampere as was done just now; he would speak of "7 milliamperes", or, more familiarly still, of "7 milliamps". A milliampere is thus seen to be a thousandth part of an ampere. Several other such convenient prefixes are in common use; the most frequent are:

Prefix.	*Meaning.*	*Symbol.*
milli-	One thousandth of	m
micro-	One millionth of	μ
kilo-	One thousand	k
mega-	One million	M

These prefixes can be put in front of any unit; one speaks commonly of milliamps., microamps., kilocycles per second, megohms, and half a dozen other such odd-sized units. "Half a megohm" comes much more trippingly off the tongue than "Five hundred thousand ohms", just as $\frac{1}{2}M\Omega$ is quicker to write than 500,000 Ω.

It must be noticed, however, that Ohm's Law refers to volts, ohms, and amperes; the indiscriminate use of odd units will lead to odd results. If a current of 5 milliamps (mA) is flowing through 15,000 ohms (Ω), the voltage across that resistance will *not* be 75,000 volts. The current must be expressed as 0·005 amp. before the correct result, 75 volts, is obtained for the magnitude of the potential difference.

12a. Kirchhoff's Law

The term *potential difference* is used above instead of E.M.F. because the voltage across the resistance is a *result* of the current, and not the cause of it. The E.M.F. driving the current must be due to a battery or other source somewhere in the circuit. Going back to our water-flow analogy; if a circulating pump forces water through a narrow pipe or a filter offering resistance to its flow, a difference in pressure is caused between the two ends of the pipe. This difference in pressure (or potential) is exactly equal to what may be called the water-motive-force, or pressure in pounds per square inch, produced by the pump. Water pressure can be measured by the height at which the liquid will rise in a vertical pipe open at the top. So it is sometimes specified as a "head" of so many feet, and a pressure difference can be regarded as a difference in level. The greater the difference in level (either feet or volts) the greater the tendency to cause a current (of water or electricity) to flow.

Whenever there is an E.M.F. in a circuit, the resulting current causes a potential difference (P.D.) that is exactly equal and opposite. The P.D. may be the result of the resistance offered by one or more sections of the circuit, or of other circuit effects that will be considered in the next chapter. So if one starts at any point in a closed circuit, reckoning the E.M.F.'s as (say) positive, and the opposing P.D.'s as negative, one must always find the result adds up to nil. If it doesn't, a mistake has been made somewhere ; just as a surveyor must have made an error if he starts off from a certain spot, reckons his increases in height above sea level as positive and his descents as negative, and after following a circular route to his starting point finds that (on paper) he has made a net ascent or descent. Obviously he can be neither higher nor lower when he arrives back at the same spot ; and in the same way the total voltage round any complete circuit, when account is taken of positive and negative, must be zero. This is a very valuable check on one's working, especially in complicated circuits. This principle, in fact, although it follows directly from Ohm's Law, is regarded as of sufficient importance to be called a Law, too, associated with the name of Kirchhoff.

The reader must be warned against a possible cause of confusion with regard to the direction of current flow. In the early days of electrical knowledge, before it was clear whether electric currents consisted of positive electricity flowing one way, or negative electricity flowing in the opposite direction, or both at once, it was agreed to guess that positive electricity flowed from one terminal of a battery (which was therefore marked +) through the external circuit to the negative or — terminal. Unfortunately it was a bad guess, for the current consists of electrons flowing from — to + terminals, or in the opposite direction to that supposed to be taken by the current. As it has become so firmly established to talk about current flowing from positive to negative terminal, this convention is adhered to in this book (note direction of arrows in Fig. 7); so it must be remembered that it is really electrons flowing in the opposite direction.

13. Electrical Power

It would be a commonplace to point out that to pump water along a horizontal pipe some small amount of *power* would be required to overcome the friction. It is equally true to say that if electricity is driven through a conductor some power is required to overcome the resistance of that conductor. A rise either in voltage (pressure), current (flow of water), or resistance (friction) will naturally increase the power necessary to maintain the flow. Since these three are related by Ohm's Law, the power needed can be expressed in terms of any two of them. Using standard symbols the power is :—$W = I^2R$, or EI, or E^2/R.

Any of these expressions can be used for calculating the power expended in a circuit, according to whether current and resistance, voltage and current, or voltage and resistance are known. Once again the units to be used are ohms, amperes, and volts, while the unit of power is the *watt*. One watt is the power expended when a current of one ampere is driven by an E.M.F. of one volt.

Take the case of an electric fire having a resistance of 20 ohms, connected to 200-volt mains. By Ohm's Law the current will be 10 amperes. The three expressions for power work out, for this case, as follows :—

$I^2R = 10^2 \times 20 = 100 \times 20 = 2{,}000$ watts.

$EI = 200 \times 10 = 2{,}000$ watts.

$E^2/R = 200^2/20 = 40{,}000/20 = 2{,}000$ watts.

When electrical energy is consumed, some other form of energy must necessarily appear in its place (Law of Conservation of Energy). In the case given it is fairly evident that the electricity consumed is converted into heat. This is equally true of any case where a current passes through a resistance, though if the dissipation of power is small, the rise in temperature may not be noticed. For example the heat developed by a 15,000 ohm resistance carrying 5mA, which only dissipates 375 milliwatts (0·375 watt) would be quite difficult to detect.

It is important to note that the watt is a unit of power, which is rate of doing work, and not of simple work or energy. A ten-horse-power engine exerts ten horse-power, no matter whether it runs for a second or a day; if it continues for an hour the work done is ten horse-power-hours. Similarly, one coulomb per second under a pressure of one volt is one watt, no matter how long it flows. If the 2,000-watt fire were left on for eight hours the power would be 2 kilowatts at any moment during that time, and the total energy expended would be 16 *kilowatt-hours*. A kilowatt-hour is the " unit " charged for in the quarterly electric-light bill.

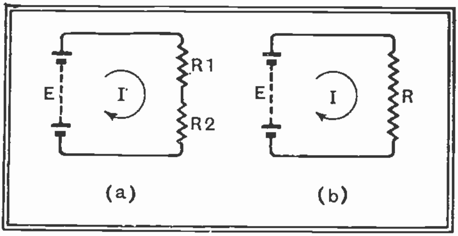


Fig. 8 : Resistances in Series. The circuit *b* is equivalent to the circuit *a*, in the sense that both take the same current from the battery E, if $R = R_1 + R_2$

14. Resistances in Series or Parallel

It is only in the simplest cases that a circuit consists of no more than a source of E.M.F. and a single resistance. A battery lighting a single lamp or a single valve-filament is one of the few practical examples. The circuits with which we shall have to deal will in most cases contain several resistances or other circuit elements, and these may be connected either *in series* or *in parallel*. Fortunately Ohm's Law can be applied to every part of a complex circuit as well as to the whole. Beginners often seem reluctant to make use of this fact, being scared by the apparent difficulty of the problem.

Two elements are said to be in series when in tracing out the path of the current we encounter them serially, one after the other. In Fig. 8 the two resistances R_1 and R_2 are connected in this way. Remembering that an electric current is an electron-flow, it will be evident that *the same current flows through both of them.*

Two elements are said to be in parallel if they are so connected in the circuit that they form two alternative paths for the current flowing between a pair of points. In Fig. 9, for example, R_1 and R_2 are alternative paths for conveying current from A to B. It will be evident, from the nature of things, that the same potential difference exists across both of them.

It does not follow that because two circuit elements have the same potential difference across them that they are necessarily to be regarded as connected in parallel. If in Fig. 8 *a* R_1 and R_2 had the same resistance, it would follow that the potential difference across R_1 would be equal to that across R_2. In spite of this fact, they are very evidently not connected in parallel; the equality of voltage across them is an accidental result of the particular relative magnitudes we have arbitrarily assigned to them, and not, as in the case of the parallel-connected resistances of Fig. 9 *a*, a necessary consequence of their mode of connection into the circuit.

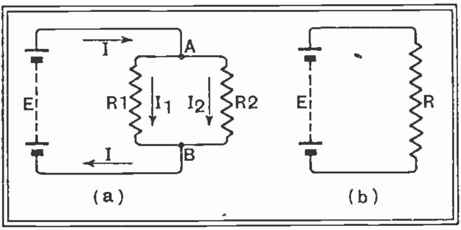


Fig. 9: Resistances in parallel. The circuit *b* is equivalent to the circuit *a* in the sense that both take the same current from the battery E, if $1/R = 1/R_1 + 1/R_2$.

In other words, R_1 and R_2 of Fig. 8 *a*, if of equal resistance have equal voltages across them, whereas R_1 and R_2, of Fig. 9 *a*, irrespective of their relative magnitudes, have *the same voltage* across them. Rather a fine distinction, perhaps, but a very vital one for the clear understanding of circuits of all kinds. Bearing this point in mind, we can shorten our definitions by saying that :—

"In Series" means *the same current.*

"In Parallel" means *the same voltage.*

15. Resistances in Series

In Fig. 8 *a* two resistances, R_1 and R_2, are shown connected in series with one another and with the battery of voltage E. To relate this circuit to the simpler ones already discussed we need to know what single resistance R (Fig. 8 *b*) can be used as a substitute for R_1 and R_2 taken together.

We know that the current in the circuit is everywhere the same ; call it I. Then the potential difference across R_1 is IR_1, and that across R_2 is IR_2 (Ohm's Law). The total voltage-drop is the sum of these two, namely, $IR_1 + IR_2$, or $I(R_1 + R_2)$, and is equal to the voltage E of the battery. In the equivalent circuit of Fig. 8 *b*, E is equal to IR, and since, to make the circuits truly equivalent, the current must be the same in both for the same battery-voltage, we see that $R = R_1 + R_2$. Generalizing from this result, we conclude that : *The total resistance of several resistances in series is equal to the sum of their individual resistances.*

16. Resistances in Parallel

Turning to the parallel-connected resistances of Fig. 9 *a*, we have the fundamental fact that they have the same voltage across them ; in this case the E.M.F. of the battery. Each of these resistances will take a current depending on its own resistance and on the E.M.F. of the battery ; the simplest case of Ohm's Law. Calling the currents respectively I_1 and I_2, we therefore know that $I_1 = E/R_1$ and $I_2 = E/R_2$. The total current drawn is the sum of the two : it is $I = E/R_1 + E/R_2 = E(1/R_1 + 1/R_2)$. In the equivalent circuit of Fig. 9 the current is E/R, which may also be written E(1/R). Since, for true equivalence between the circuits, the current must be the same for the same battery voltage, we see that $1/R = 1/R_1 + 1/R_2$. Generalizing from this result, we may conclude that : *If several resistances are connected in parallel the sum of the reciprocals of their individual resistances is equal to the reciprocal of their total resistance.*

If the resistances of Fig. 9 *a* were 100 and 200 ohms, the single resistance R that, connected in their place, would draw the same current is given by $1/R = 1/100 + 1/200 = 0{\cdot}01 + 0{\cdot}005 = 0{\cdot}015$. Hence, $R = 1/0{\cdot}015 = 66{\cdot}67$ ohms. This could be checked by summing the individual currents through 100 and 200 ohms, and comparing the total with the current taken from the same voltage-source by 66·67 ohms. In both cases the result is 0·015 ampere per volt of battery.

Summing up, we have the two rules which, expressed in symbolic form, are :—

1. Series Connection. $R = R_1 + R_2 + R_3 + R_4 + \ldots\ldots$
2. Parallel Connection. $1/R = 1/R_1 + 1/R_2 + 1/R_3 + \ldots\ldots$

For two resistances in parallel, a more easily calculated form of the same rule is $R = R_1R_2/(R_1 + R_2)$. More than two resistances in parallel can be dealt with by first working out the equivalent of two of them, and then combining this with the next ; and so on.

17. Series-Parallel Combinations

These rules can be extended to cover quite a complicated network of resistances. In such cases the algebra required, though perfectly simple, is inclined to get very long-winded if an attempt is made to work out a general formula ; we will therefore content ourselves with one example, worked out numerically. The example will be the circuit of Fig. 10 ; we will find the total current flowing, the equivalent resistance of the whole circuit, and the voltage and current of every resistor individually.

The bunch R_2, R_3, R_4, R_5 is obviously going to be our stumbling block, so we will begin by simplifying it. In doing this it is always necessary to work from the inside outwards. Writing R_{23} to symbolize the combined resistance of R_2 and R_3 taken together, we know that $R_{23} = R_2R_3/(R_2 + R_3) = 200 \times 500/700 = 142{\cdot}8$ ohms. This gives us the simplified circuit of Fig. 11 *a*. If R_{23} and R_4 were one resistance, they and R_5 in parallel would make another simple case, so we proceed to combine R_{23} and R_4 to make R_{234}.

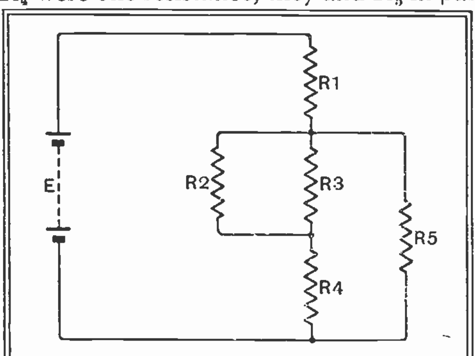


Fig. 10 : A complicated network of resistances. The current through and voltage across each can be computed with the aid of the rules already discussed. R1 = 100 ohms, R2 = 200 ohms, R3 = 500 ohms, R4 = 150 ohms, R5 = 1000 ohms, E = 40 volts

$R_{234} = R_2 + R_4 = 142{\cdot}8 + 150 = 292{\cdot}8$ ohms. Now we have the circuit of Fig. 11*b*. Combining R_{234} and R_5, $R_{2345} = 292{\cdot}8 \times 1000/1292{\cdot}8 = 226{\cdot}5$ ohms. This brings us within sight of the end ; Fig. 11 *c* shows us that the total resistance of the network now is simply the sum of the two remaining resistances ; that is, R of Fig. 11 *d* is $R_{2345} + R_1 = 226 \cdot 5 + 100 = 326 \cdot 5$ ohms.

From the point of view of current drawn from the 40-volt source the whole system of Fig. 10 is equivalent to a single resistance of this value. The current taken from the battery will therefore be $40/326 \cdot 5 = 0 \cdot 1225$ amp. $= 122 \cdot 5$ mA.

To find the current through each resistor individually now merely means the application of Ohm's Law to some of our previous results. Since R_1 carries the whole current of 122·5 mA, the potential difference across it will be $100 \times 0 \cdot 1225 = 12 \cdot 25$ volts. R_{2345} also carries the whole current (11 *c*) ; the p.d. across it will again be the product of resistance and current, in this case $226 \cdot 5 \times 0 \cdot 1225 = 27 \cdot 75$ volts. This same voltage also exists, as comparison of the various diagrams will show, across the whole complex system R_2 R_3 R_4 R_5 in Fig. 10. Across R_5 there lies the whole of this voltage ; the current through this resistor will therefore be $27 \cdot 75/1000$ amp. $= 27 \cdot 75$ mA.

The same p.d. across R_{234} of Fig. 11 *b*, or across the system R_2 R_3 R_4 of Fig. 10, will drive a current of $27 \cdot 75/292 \cdot 8 = 94 \cdot 75$ mA through this branch. The whole of this flows through R_4 (11 *a*), the voltage across which will accordingly be $150 \times 0 \cdot 09475 = 14 \cdot 21$ v. Similarly, the p.d. across R_{23} in Fig. 11 *a*, or across both R_2 and R_3 in Fig. 10, will be $0 \cdot 09475 \times 142 \cdot 8 = 13 \cdot 54$ volts, from which we find that the currents through R_2 and R_3 will be respectively $13 \cdot 54/200$ and $13 \cdot 54/500$ amp., or 67·68 and 27·07 mA, making up the required total of 94·75 mA for this branch.

This gives a complete analysis of the entire circuit ; we can now collect our scattered results in the form of the following table :—

RESULTS OF SOLVING FIG. 10

Resistance	Current [milliamps.]	Voltage [volts]	Power [watts]
R_1	122·5	12·25	1·501
R_2	67·68	13·34	0·916
R_3	27·07	13·54	0·367
R_4	94·75	14·21	1·346
R_5	27·75	27·75	0·771

It is instructive to apply Kirchhoff's Law to the several closed circuits that are included in this network. Consider that formed by E, R_1, and R_5. If we take the route clockwise we go "uphill" through the battery, becoming more positive by 40 volts. Coming down through R_1 we move from positive to negative so add -- 12·25 volts, and, through R_5, − 27·75 volts, reaching the starting point again. Check : $40 - 12{\cdot}25 - 27{\cdot}75 = 0$.

Taking another clockwise route via R_3, R_5, and R_4, we get $13{\cdot}54 - 27{\cdot}75 + 14{\cdot}21 = 0$. And so on for any closed loop.

It should be noted that by using suitable resistors any potential intermediate between those given by the terminals of the battery can be obtained. For instance, if the lower and upper ends of the battery in Fig. 10 are regarded as 0 and + 40 (they could equally be reckoned as − 40 and 0, or − 10 and + 30, with respect to any selected level of voltage), the potential of the junction between R_3 and R_4 is 14·21 volts. The arrangement is therefore called a *potential divider*, and is often employed for obtaining a desired potential not given directly by the terminals of the source. If a sliding connection is provided on a resistor, to give a continuously variable potential, it is generally known—though not always quite justifiably—as a *potentiometer*.

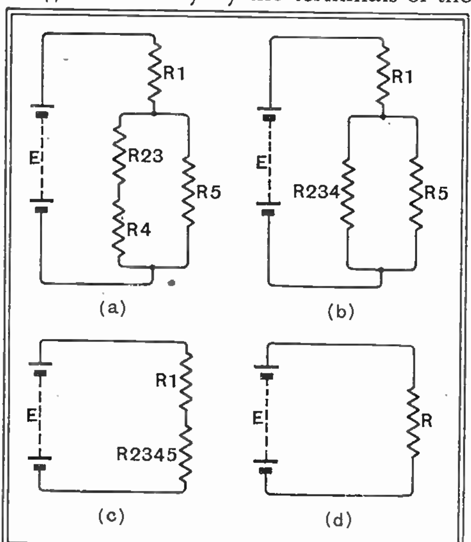


Fig. 11 : Successive stages in simplifying the circuit of Fig. 10. R_{23} stands for the single resistances equivalent to R_2 and R_3 ; R_{234} to that equivalent to R_2, R_3 and R_4. and so on. R represents the whole system

CHAPTER 3

INDUCTANCE AND CAPACITANCE

18. Magnets and Electromagnets

IF a piece of paper is laid on a straight " bar " magnet, and iron filings are sprinkled on this paper, they are found to arrange themselves in some such pattern as that indicated in Fig. 12. These lines show the paths along which the attraction of the magnet exerts itself, and so are called *lines of magnetic force*. As a whole, they map out the *magnetic field*, which is the region in which the effect of the magnet is felt. (Compare the *electric* field in the previous chapter.)

An electric charge on a body represents, as we have seen (Sec. 8), a certain amount of stored energy ; a magnetic field contains stored energy in another form. This energy is limited in amount, and can only be made use of at the cost of destroying the field, just as the energy of a charged body can only be liberated by allowing it to drive a current through a circuit, and so dissipating the charge.

Neither a magnet nor an electrified body is giving out or demanding a continuous flow of energy merely in maintaining its field, any more than a tank of water at the top of the house. In an electro-magnet, which consists, as Fig. 13 shows, of a coil of wire surrounding an iron core, it is found that the magnetic effect is set up when the current is turned on,

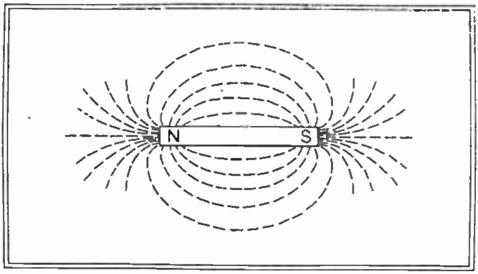


Fig. 12 : The lines of magnetic force round a permanent magnet N.S. These lines mark out the magnetic field surrounding the magnet

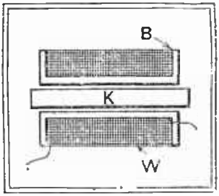


Fig. 13 : Section through an electro-magnet. K, iron core ; B, bobbin fitting over K ; W, winding of insulated copper wire

remains as long as the current through the coil continues, and vanishes when the current stops. The energy necessary to create this field has to come from somewhere—there being no other source, it must come from the current. This means that while the field is being built up the battery has to drive current against an opposition greater than that due to the mere resistance of the wire, so that *while the field is growing*, the electro-magnet behaves rather as though it contained extra resistance. But once the field is set up, no energy is required to maintain it. The current through the magnet then becomes, and remains, exactly what one would predict from the E.M.F. of the battery and the pure resistance of the wire of the coil; the magnetic field plays no part in determining the magnitude of the current once it has settled down to a steady value.

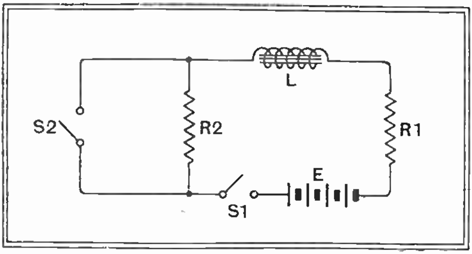


Fig. 14 : Current in inductive circuit. L, electro-magnet (inductance): R_1, represents resistance of L ; R_2, high resistance switched into circuit by opening S_2 ; S_1, main circuit switch

It is a little difficult to visualize what happens on switching off the current, because of the rather uncertain nature of a switch, which may spark across the contacts. Instead, we will imagine that the current is reduced to one-thousandth of its original steady value by opening a switch connected across a resistance of high ohmic value, as suggested in Fig. 14. When the current drops the magnetic field will collapse with it, and experiments show that the stored energy that it contained makes itself felt as an attempt towards maintaining the full current. Naturally, since the

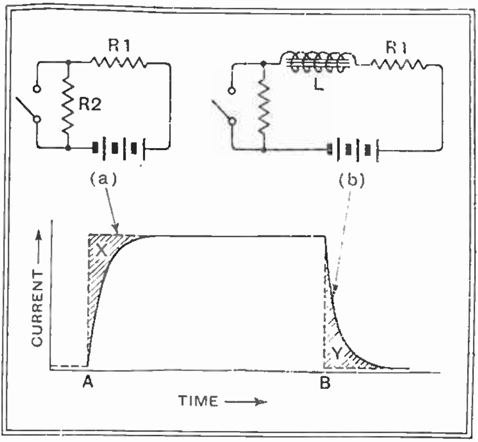


Fig. 15 : The rise and decay of current in a circuit which is (*a*) resistive (dotted line) ; (*b*) inductive (full line). In the presence of inductance the current takes time to attain a new value

energy of the field is limited, this attempt will not succeed. The effect is that for an instant the current is higher than would be calculated from Ohm's Law by taking into account the E.M.F. of the battery and the new, high value of the resistance of the circuit. It is in this way that the energy originally taken for building up the field is returned to the circuit when the field collapses.

19. Inductance

The foregoing paragraphs can be summarized by saying that the effect of the field is to check the current when it is rising, and to maintain it when falling ; in brief, *to oppose any change in the current.* So long as the current is steady the presence of the field does not affect it.

These points are shown graphically by the curves of Fig. 15. In these, time is plotted from left to right and current upwards ; the dotted curves refer to a circuit containing only an ordinary resistance, while the full-line curves refer to a circuit containing an electro-magnet. In the circuit comprising only E.M.F. and resistance (Inset *a*) the current rises instantaneously to full value at the exact instant of switching out the high resistance R_2 (A on the curve) as shown by the dotted line. At B, the instant of switching in the high resistance, it falls instantaneously to its new, very low, value. In the circuit including the electro-magnet (Inset *b*) the current rises more slowly, requiring, as the full line shows, an appreciable time to reach its full value. At B the retarding effect of the magnetic field, now returning energy to the circuit, makes the fall in current slow, the change again taking place according to the full line. The shaded area X represents the energy used in building up the field, while the area Y represents the energy returned in prolonging the current, when the field collapsed. The two areas are equal.

This property, by which an electrical circuit offers opposition to the *change* of a current flowing in it, is called *inductance*. It must always exist in any practical circuit, for there is a magnetic field round even a straight wire so long as it carries a current. In practice, the effect is seldom noticeable until the wire is made into a coil, so that the fields due to the different parts of the circuit can reinforce one another. The presence of an iron core enhances the effect immensely, since the lines of force can pass far more readily through iron than through air.

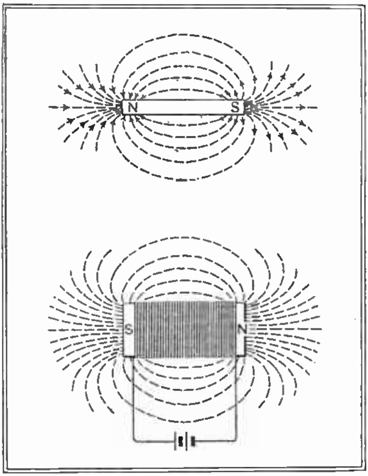


Fig. 16 : Conventional representations, in terms of "lines of magnetic force" of the fields round a magnet and a coil of wire carrying a current

Inductance is measured in *henrys* (Symbol H). This unit is defined by the condition that if the current flowing in a circuit changes by one ampere when a potential difference of one volt is applied for one second, the circuit has an inductance of one henry.

It is conventional to represent a magnet, as in Fig. 16, as being the source of lines of "magnetic force", which leave the magnet by the North Pole and re-enter it by the South.

20. Lines of Magnetic Force

The lines trace out the path along which a north magnetic pole, if free to move, would be impelled by the field, and are as real, or as unreal, as the parallels of latitude on the map. The field is at its most intense in the neighbourhood of the poles, but theoretically it extends infinitely in all directions, dying away rapidly in intensity as we retreat from the magnet.

By sending a steady current through a coil of wire this becomes, as we have seen, a magnet, and remains so as long as the current is maintained. The magnitude of the field round an electro-magnet depends upon the number of the turns and on the current through them. Ten turns carrying one ampere gives rise to the same field as a hundred turns carrying one-tenth of an ampere; ten *ampere-turns* are available in either case to set up the field.

21. Interacting Magnetic Fields

Everybody knows that a compass-needle will point to the north. The needle itself is a magnet, and turns because its own field interacts with the magnetic field of the earth. Put differently, the north magnetic pole of the earth attracts the north-seeking pole of the magnet while the earth's south pole attracts its south-seeking pole. This is the only known case of two "north poles" or two

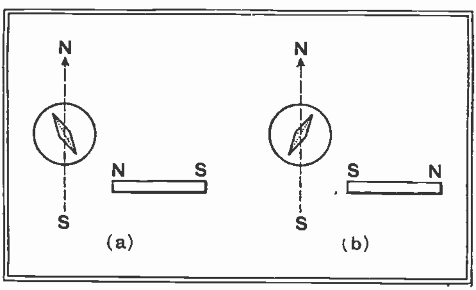


Fig. 17: Deflection of a compass needle by a magnet, proving that like poles repel and unlike poles attract. The N and S poles of the bar magnet may be found by suspending it like a compass needle and marking as N that pole which turns to the north

"south poles" attracting one another, and is simply due to convenient, but muddled, nomenclature; the earth's "north pole" has the same polarity as the "south pole" (more correctly, south-*seeking* pole) of a magnet. By bringing the two poles of a bar magnet, in turn, towards a compass needle it is very easy to demonstrate that, as in Fig. 17, like poles repel one another and unlike poles attract. This reminds one, again, of the way electric charges behave.

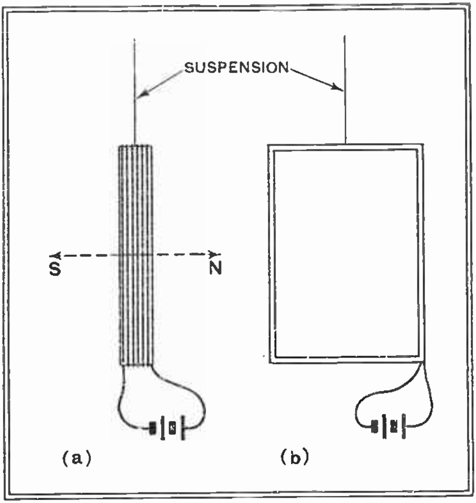


Fig. 18: A coil, free to rotate, sets itself in the orientation indicated, when a current is passed through the winding

Now suppose we hang a coil in such a way that it is free to rotate about a vertical axis, as suggested in Fig. 18. So long as no current is passed through the coil it will evince no tendency to set itself in any particular direction, but if a battery is connected to it the flow of current will transform the coil into a magnet. Like the compass needle, it will then indicate the north, turning itself so that the plane in which the turns of the coil lie is east and west, the axis of the coil pointing north. If the current is now reversed the coil will turn through 180 degrees, showing that what was the north pole of the coil is now, with the current flowing the opposite way, the south.

The earth's field is weak, so that the force operating to turn the coil is small; when it is desired to make mechanical use of the magnetic effect of the current in a coil it is usual to provide an artificial field of the highest possible intensity by placing a powerful magnet as close to the coil as possible.

22. Measuring Currents and Voltages

The tendency to turn exhibited by a coil carrying a a current depends upon the intensity of the magnetic field in which the coil lies, and upon the ampere-turns available to provide the coil's own field. In a constant external field, a coil of a fixed number of turns is rotated by a force depending only on the current passed through it; if the coil, in turning, is compelled to wind up a light spring, the degree of rotation will be a measure of the current causing it.

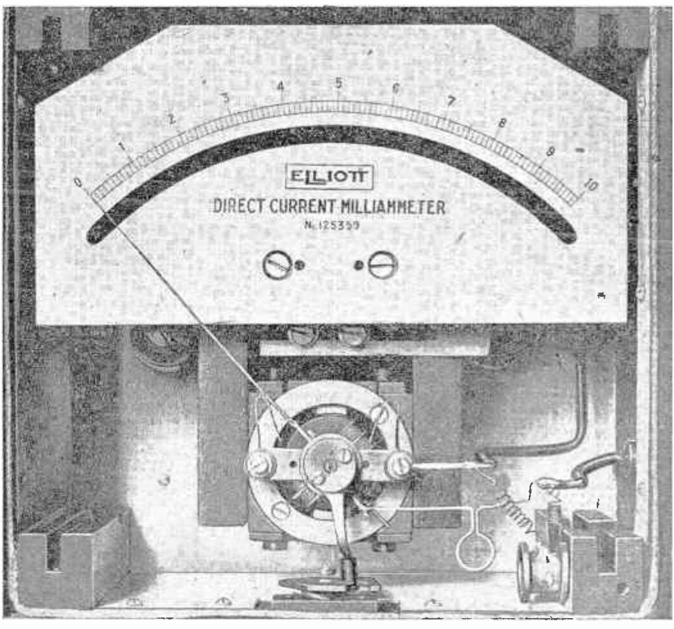


Fig. 19 : Milliammeter with front removed. The coil lies behind the pivot between the poles of the magnet. When moved by a current, it sweeps the pointer over the scale

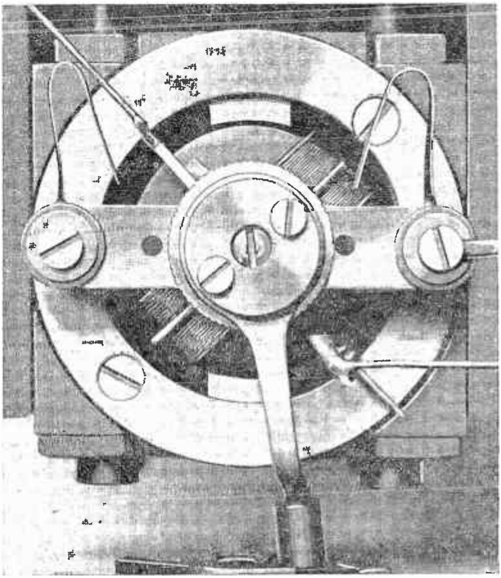


Fig. 20: Enlarged view of pivot of milliammeter. The coil can be seen here; it moves at right-angles to the pointer. Note balancing-weight on the latter

This is the principle of the *moving-coil meter.* The current to be measured is caused to pass through a coil of wire suspended on light bearings between the poles of a permanent magnet (see Figs. 19 and 20). The magnetic field set up by the passage of the current through the coil is acted upon by the field of the magnet, the two being so disposed that the resulting mechanical force tends to rotate the coil. Except for the restoring torque of a light spring, this is free to turn, carrying with it a pointer which moves across a scale calibrated in amperes or milliamperes. In the former case the instrument is called an *ammeter*, in the latter a *milliammeter.*

In order that the current in a circuit may not be appreciably altered when the meter is inserted to read it, an ammeter or milliammeter always has a low resistance.

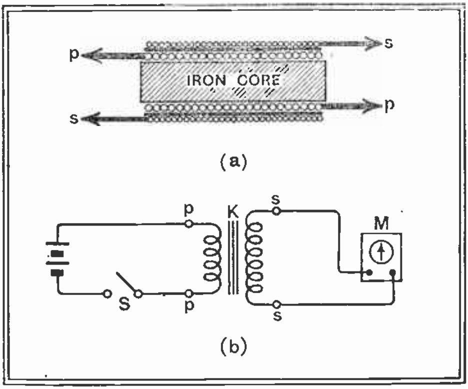
A *voltmeter*, which is scaled to read volts directly, is in reality a milliammeter in series with a fixed resistance of high value. If scaled in milliamperes, all readings would have to be multiplied by the value of this resistance (since E = IR) to find the corresponding voltages ; in a proper voltmeter this multiplication is done once and for all by the maker of the instrument when he engraves the scale, which therefore reads volts directly.

Although very much better methods exist, and are used in laboratory work, it is usual to measure resistance in everyday wireless practice by observing independently the current through and the voltage across, the resistance to be measured. Ohm's Law then gives the resistance by simple division of voltage by current.

23. Induced Voltages

The energy stored in a magnetic field can be converted into an electric current in other ways than that illustrated in Fig. 15. For example, we can replace the simple coil of that figure by two coils, one wound over the other in the way suggested in Fig. 21 *a*, where the ends of one winding are marked *p* and the ends of the other marked *s*. In Fig. 21 *b* these two coils are shown connected into a circuit designed to demonstrate the existence of *induced voltages*. The *primary* winding *p* is connected, through a switch S, to a battery. The ends *s* of the *secondary* winding are connected to a centre-zero milliammeter M, in which the pointer, when at rest, lies in the middle of the scale, deflecting to right or left according to the direction of the current sent through it.

On closing S the current through the primary rises to a value which, when the steady state is reached (Fig. 15) must depend solely on the voltage of the battery and the total resistance in the primary circuit. This current, and the magnetic field it evokes, are thus quite unaffected by the presence of the secondary winding. But at the moment when the current is switched on, the milliammeter M is seen to give a violent kick, returning immediately to zero. This shows that a current has flowed in the secondary during the period in which the current in the primary, and hence the magnetic field round both coils, *was building up*. On opening the



switch, M registers the momentary flow of a current in the opposite direction to that observed on closing it, while the current in the primary can be shown to drop practically instantaneously to zero.

Fig. 21 : (a) Two separate coils wound over a common iron core. p p primary : s s secondary. (b) Conventionally-drawn circuit showing the coils of *a* connected so as to demonstrate that changes of current in p p cause induced voltages in s s. The iron core is represented at K

This latter observation shows that the energy of the magnetic field has been used up in driving a current through the *secondary* instead of in maintaining, as in Fig. 15, the current in the primary.

For these currents to be set up in the secondary, some driving voltage must have been present there. This voltage must be derived from the change in the magnetic field surrounding the secondary, since there is no communication between the two windings other than that established through the field surrounding both. We conclude, therefore, that a voltage is induced in a coil whenever there is a *change* in the magnetic field surrounding it. This can be checked by removing the primary winding and the iron core, and pushing the end of a bar magnet in and out of the remaining coil. M is seen to deflect momentarily every time the magnet is moved in or out, the direction of the current induced by pushing a north pole in or drawing a south pole out being opposite to that induced by pulling a north pole out or pushing a south pole in.

The energy of the current is not derived from the magnet, which retains its field unimpaired. Instead, it comes from the mechanical effort necessary to insert or withdraw the magnet against the resistance set up by the interaction of the field due to the current with that of the magnet.

24. Inductance Redefined

It will be clear that in Fig. 15, too, we have a coil surrounded by a varying magnetic field, only in this case the coil affected by the field is also the coil used to produce it. We can therefore attribute the slow rise and slow fall of the current to a voltage, induced by the changing field, in opposition to the change in voltage across the coil produced by opening or closing the switch. This leads us to an alternative, but equally satisfactory, definition of the unit of inductance; we may say that: A coil has an inductance of one henry when a change in the current at the rate of one ampere per second produces across it an induced E.M.F. of one volt.

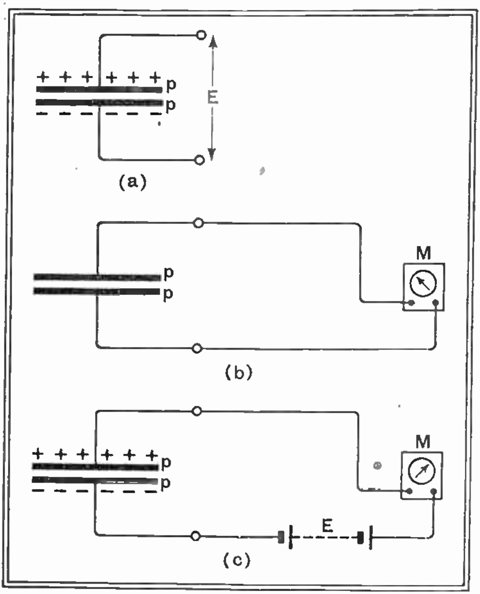


Fig. 22: (a) A condenser charged to a difference of potential E volts: two metal plates pp separated by air. (b) On joining the plates through a centre-zero milliammeter M, a momentary current flows, and the difference of potential between the plates vanishes. The condenser is now *discharged*. (c) On inserting a battery of potential E volts, a momentary current again flows, now in the opposite direction, and the plates take on their original charges as at *a*. The condenser is now *charged* again

25. Mutual Inductance

We can go further, and define on similar lines the effect that one coil, through its magnetic field, exerts on another. In such a case we say that: The *mutual inductance* between two coils is one henry when a change in current at the rate of one ampere per second in the one induces an E.M.F. of one volt in the other.

26. Capacitance

Suppose that, as in Fig. 22 *a*, we have two metal plates p, p, separated by a thin layer of air. We know (Sec. 8) that if a lot of electrons have by some means been taken from the upper plate and put on the lower, the two plates are oppositely charged, as indicated by the + and – signs, and that their attraction for one another appears both as a mechanical attraction of the two plates for one another, and as an electrical E.M.F. tending to drive electrons from the lower to the upper plate. On joining them by a conductor, therefore, a current of electrons flows in that direction (conventionally spoken of as an electric current in the opposite direction—positive to negative; see Sec. 12A), indicated by the central-zero milliammeter at *b*. If the conductor is of low resistance the current is very short-lived, however, because the distribution of electrons is quickly equalised and the E.M.F. consequently ceases to exist.

If now we were to open the circuit and insert a battery of voltage E (i.e., having between its terminals an E.M.F. equal to that originally on the plates at *a*), this process will be reversed, and a current equal to that shown at *b*, but opposite in direction, will flow round the circuit until the plates are again charged to the difference of potential E. The latter current, driven by the battery, is known as a *charging current*, while the former is called a *discharging current*. The assemblage of plates, as a whole, is known as a *condenser*.

Beginners often have difficulty in visualising a current in a circuit where there is a complete break in the conducting path, namely, a condenser. The distinction between it and a current in a conductor is expressed by saying that the current flows *through* a conductor but *into* a condenser. Turning once more to our waterpipe analogy, it is true that if there is a complete and unyielding block in the pipe no water can be made to flow. But if the block takes the form of a sheet of rubber it is possible to force a limited quantity of water from one side to the other, although actually no water flows *through*. The rubber diaphragm, in being distended in this way, causes a back pressure that will drive the water back again if the original pressure is relaxed. Incidentally, also, an excessive pressure may rupture the diaphragm, converting the elastically blocked pipe into an unobstructed pipe and allowing a continuous flow. In the same way, if an excessive E.M.F. is applied to a condenser it can be broken down and converted into a conductor. For a more electrical picture of a charging and discharging current see Sec. 28.

27. The Unit of Capacitance

If the area of the two plates in Fig. 22 were doubled, it is not difficult to see that twice as many electrons would be required to charge them to the difference of potential E If the separation of the plates were halved, only half the total P.D., E, would be needed to maintain the same intensity of field in the gap ; therefore if the *whole* of the original charging voltage E were applied, double the number of electrons would be transferred from one plate to the other. The electron-storing ability of the condenser, known technically as its *capacitance*, can evidently be measured in terms of the quantity of electricity (in coulombs) stored for each volt applied in charging it ; thus we may express the capacitance C by the formula $C = Q/E$. Evidently the capacitance will be one unit—one *farad*—if one volt drives one coulomb of electricity into the condenser.

It so happens that a condenser of capacitance one farad would completely fill the average small room. The wireless engineer works in smaller units, and uses microfarads (μF, or mfd.*) or even micromicrofarads ($\mu\mu$F or mmfd.*). which are of a more convenient size both electrically and mechanically. It is always necessary, of course, to use the full farad for theoretical calculations.

* *Warning*. "mfd." *ought* to mean "millifarad", but it is never used with this correct meaning. This abbreviation seems to have been made merely phonetically.

28. Dielectrics

In a condenser consisting of two plates separated by air the capacitance may be, perhaps, 0·001 μF., or 1,000 $\mu\mu$F. This can be measured by observing the quantity of electricity (10^{-9}coulombs) necessary to charge it to one volt. If we now fill up all the space between the plates with a different insulating material, say, mica, we shall find, on remeasuring the capacitance, that it has increased to several times its original value.

This effect can best be understood if we consider the effect of the electric field between the plates upon the insulating material, or *dielectric*, that we have placed there. Being an insulator, this material does not contain electrons free enough to move from atom to atom under the urge of an electric field, for that is the characteristic of a conductor. But the electrons can move, to a limited extent, within the limits of their atoms, and this movement stops when the elastic forces within the atoms, which tend to return the electrons to their normal places, exactly counterbalance the driving force of the field applied. The energy thus stored in the dielectric is additional to that stored in the space between the plates; to charge the condenser to one volt therefore requires more electricity when there is a dielectric between the plates than when they are separated only by empty space, and the capacitance of the condenser is correspondingly raised by its presence.

The ratio of the capacitance of the condenser with the dielectric present to the capacitance when there is only air (more strictly, a vacuum) between the plates is known as the *dielectric constant* or *permittivity* of the material. (Sometimes called "specific inductive capacity", abbreviated to "S.I.C." This is an older term, now going out of use.)

The action of the dielectric in storing energy is exactly analogous to that of a spring put under tension. The extension of the spring depends on the strength of the pull, the spring can be broken by the application of sufficient force, and unless broken will return sharply to its original length when released. Similarly, the movement of electrons in the dielectric is greater for greater applied voltages, the insulation can be broken down, allowing a continuous current to flow, if the voltage is high enough, or the electrons revert to their usual places, thus producing a momentary current in the reverse direction, if the voltage is removed.

29. Practical Forms of Condenser

Any two plates separated from one another form a condenser, but varying modes of construction are adopted for varying purposes. A variable tuning condenser, of capacitance up to some 0·0005 mfd., consists of two sets of metallic vanes which can be progressively interleaved with one another to obtain any desired capacitance up to the maximum available. "Fixed condensers" consist of a number of metal plates interleaved with thin sheets of mica or, if a large capacitance is required, of two long strips of metal foil separated by waxed paper and rolled up into a compact block. Where the passage of a small amount of direct current from plate to plate does not matter, electrolytic condensers are used to give a very high capacitance in a small space at moderate cost.

The capacitances most used run from 0·0001 μF to 0·01 μF with mica insulation, 0·01 to 4 μF with paper insulation, and 4 to 50 μF in the electrolytic type. Small capacitance condensers often use ceramic material, such as steatite, as the dielectric.

CHAPTER 4

ALTERNATING CURRENTS

30. Generating a High-Frequency Alternating Current

IN wireless we are habitually making use of *alternating* currents; that is to say, currents that are constantly reversing their direction of flow. The currents produced by the action of the microphone, and corresponding to the air waves set up by speech or music, are of this type. So are the currents used to carry the programme across long distances. And so, too, are those used to light and heat our homes and provide the power (in place of the more expensive and troublesome batteries) that work our wireless sets. The only basic difference between all these is the number of double-reversals the current makes in a second; in a word, the frequency (Sec. 5).

There is no hard-and-fast division between one lot of frequencies and another; but those below 100 cycles per second (c/s) are used for power (the standard in this country is 50 c/s); those from 25 to 10,000 or so are audible, and so are classed as audio frequencies (A.F.), while those above about 20,000 are more or less suitable for carrying signals across space, and are known as radio frequencies (R.F.). Certain of these, notably 550,000 to 1,550,000 are allocated for broadcasting.

The alternating currents of power frequency are generated by rotating machinery at the power station; but this method is not so suitable for radio frequencies. It is preferable to make use of the properties of resistance, inductance, and capacitance. In Fig. 23 there is shown a coil L connected in parallel with a condenser C, making a closed circuit. A battery B is connected across the whole, the battery-circuit being made and broken as required by the switch K. In addition, there is a resistance *r*, compared with which the resistance of the coil will be regarded as negligible, since L is to be thought of as wound with heavy-gauge wire.

If K is closed, completing the battery-circuit, a current whose magnitude will primarily be determined by the voltage of the battery and the value of *r* will flow through the coil L (after the preliminary effects due to the current growth have died down). This current will create a magnetic field round L, so that the state of affairs while the current is flowing steadily may be represented by Fig. 24, which shows the field in dotted lines. As the resistance of L has been assumed negligible, there is negligible voltage across it, and negligible charge in C.

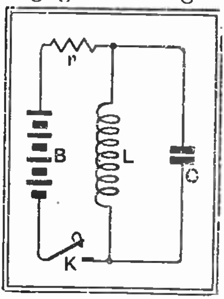


Fig. 23 : With the aid of this simple circuit the nature of high-frequency currents can be elucidated

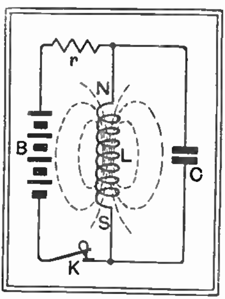


Fig. 24 : The circuit of Fig. 23 with the switch K closed. Note the magnetic field round L

At the instant when the current is interrupted again by opening K, the magnetic field contains stored energy. While the field is in process of collapsing it tends to maintain through L a current in the same direction as that which has just been interrupted. This current has nowhere to flow except into the condenser C, which thereby becomes charged as indicated in Fig. 25. The absence of magnetic lines round the coil in this figure indicates that the state of affairs represented corresponds to the moment of cessation of current, the whole energy of the magnetic field having been transferred in the form of charge (displaced electrons) to the condenser. Clearly this is not a stable condition because there is nothing to keep the condenser charged ;

the condenser will now discharge through L, driving through it a current in a direction opposite to that of the current originally provided by the battery, and building up anew the magnetic field, though now with its north and south poles interchanged. When the condenser is completely discharged, as indicated in Fig. 26, the current in the coil is at its greatest value, and the energy drawn out again from C is once more in the form of a magnetic field round L. Just as before, the field now takes over the duty of driving the current in the same direction, until it has totally collapsed, thus transferring the energy once more to C in the form of a charge, but opposite in polarity to that shown in Fig. 25.

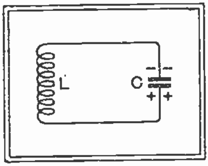


Fig. 25: The collapse of the magnetic field round L has caused a current which has charged C

If it were not that the circuit LC contains resistance in one form or another—for example, the resistance of the wire with which L is wound—the coil and condenser would continue for ever to play battledore and shuttlecock with the original supply of energy, and the current would never cease oscillating in and out of the condenser, travelling backwards and forwards through the coil for all time. In practice, of course, the resistance of the coil will dissipate the energy available in the form of heat after comparatively few interchanges.

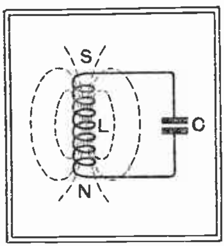


Fig. 26: The discharge of C has driven a current through L, creating a magnetic field opposite in polarity to that of Fig 24

31. The Sine-wave

Later (Sec. 108), we shall see that it is possible to supply energy to a circuit of this kind in such a way as to overcome this loss and maintain the oscillations at a steady strength. It is customary to represent such maintained oscillations of current by a diagram of the type shown in Fig. 27.

In this figure, lapse of time is indicated by distance from the left of the diagram, while magnitude and direction of current are shown by vertical distance from the line OBD, the height of which indicates zero current. Distance *above* the line is regarded as positive, indicating current flowing in one direction; while distance below the line is negative, indicating current flowing in the opposite direction. A dot anywhere on the surface of the paper would thus mean a certain current at a certain time, while a series of dots could be obtained by following the variation of a changing current from instant to instant. If we were to follow the current in the oscillating circuit LC of Figs. 23 to 26, making the assumption that the oscillation is maintained, we could mark in dots corresponding to a number of instantaneous measurements, and then join up the dots with a continuous curve to fill in the gaps. The result would be a curve like that of Fig. 27.

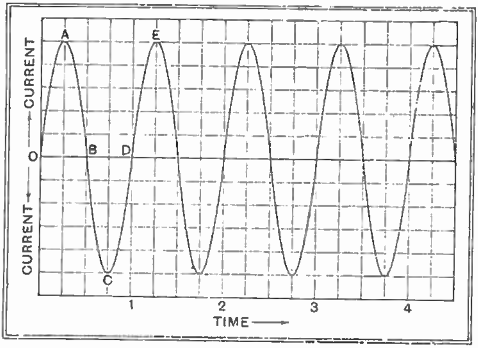


Fig. 27 : A curve showing the variation of the current in an oscillating circuit with time

On the diagram, A represents the moment of maximum current, when the magnetic field of the coil is at its greatest. From A to B the field is collapsing and the current is decreasing, until at B the current is zero and the condenser fully charged. At B the current reverses as the condenser begins to discharge again, the reversal of direction being shown by the fact that the curve now goes below the zero-line OBD. At C the current has again reached its maximum value in the reverse direction, while the charge on the condenser is gone. So the process continues until E is reached, when conditions are an exact duplicate of those existing one cycle earlier at A.

The curve is thus a faithful record of the flow of current in the circuit, but it must not be regarded as depicting the actual shape of anything, except in a purely mathematical sense. It conveys merely that the current varies with time in the manner shown, flowing first in one direction and then in the other. The steepness of the curve at any point indicates the rate at which the current is growing or decaying at that instant. The same curve can also be used to to indicate the voltage across the condenser C, which rises and falls according to the same law as the current. It is interesting to note that if the vertical position of a freely-swinging pendulum be regarded as corresponding to zero voltage (no tendency to fall) and if displacements to left and right be represented as above and below the line OBD, the same curve can be used to express its motion when swinging. An accurate mental picture of the flow of current in an oscillating circuit can therefore be acquired by watching a pendulum and allowing the imagination a little disciplined freedom.

It can be shown mathematically that curves representing any types of oscillations are all combinations of one basic curve, called the sine curve or wave (adjective : sinusoidal). The simplest and purest oscillation or alternating current, therefore, is one which varies sinusoidally, as shown in Fig. 27. We shall see later that alternating current or voltage curves of more complicated appearance are formed by combining two or more sine waves of different frequency.

32. R.M.S. and Peak Values

In Fig. 28, G is assumed to be a generator of alternating voltage, driving an alternating current through the resistance R. We have already seen (Sec. 19) that a resistance offers opposition to the flow of current, but is indifferent to changes in that current. Put differently, the sudden application or withdrawal of

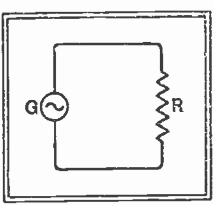


Fig. 28 : An alternating voltage is applied by the generator G (nature unspecified) to the resistance R

a voltage produces *instantaneously*, a current of the magnitude that Ohm's Law predicts. If therefore the full-line curve of Fig. 29 is taken to represent the variations of the voltage of G with time, the current will follow a curve identical in shape, though different in scale, for the current at any instant will be equal to the voltage divided by the resistance. The current curve is shown dotted.

The alternating current supplied for house lighting has, as a general rule, a frequency of 50 cycles per second (50 c/s). In terms of Fig. 27, this means that the time-scale is such that the distance A to E or O to D represents one-fiftieth of a second. Each second thus contains 50 current-pulses in each direction, so that if the temperature of its filament could change quickly enough a lamp connected to such mains would not emit a continuous light, but a series of separate flashes succeeding one another at the rate of 100 per second.

How are electric mains that behave in this fashion to be rated ?—that is, what are we going to mean when we speak of " 200-volt 50-cycle mains " seeing that the actual voltage is changing all the time, and is sometimes zero ?

The convention that has been arrived at is based on comparison with direct-current (D.C.) mains. It is obviously going to be a great convenience for everybody if a lamp or a fire intended for a 200-volt D.C. system should be equally suited to alternating mains of the same nominal voltage. This condition will be fulfilled if the *average power* taken by the lamp or fire is the same for both types of current, for then the filament will reach the same temperature and the cost of running will be the same in the two cases.

In Fig. 29 both voltage and current are shown for a resistive circuit. At any instant the power being consumed is given by the product of voltage and current. At the instant corresponding to P both are at their maximum value, and the power dissipated is also at its highest. At Q voltage and current are both zero ; so also is the power. The average power must lie somewhere between these extremes.

If by " 200-volt mains " we mean a supply whose *peak* voltage (point P) is 200 volts the maximum instantaneous power drawn by a lamp or fire would be the same as the power it would take from 200 volt D.C. mains, but the *average* power would be less. To raise this to the figure for D.C. mains, the peak voltage of the A.C. supply will evidently have to rise well above the rated nominal voltage.

It can be shown mathematically that for a curve ("sine-wave") of the form shown in Fig. 29 the average power is exactly half of that corresponding to the instant P; it is therefore half EI where E is the peak voltage and I the peak current. We need, therefore, to raise the peak voltage sufficiently far above the nominal voltage to double the peak power.

We cannot do this by simply doubling E, because this also doubles I, making the peak power four times as great. To double the power we have to increase the peak voltage $\sqrt{2}$ times, which simultaneously causes the peak current to increase $\sqrt{2}$ times. The increase of power is then to $\sqrt{2} \times \sqrt{2}$ times, or double, its original value, which is what is required.

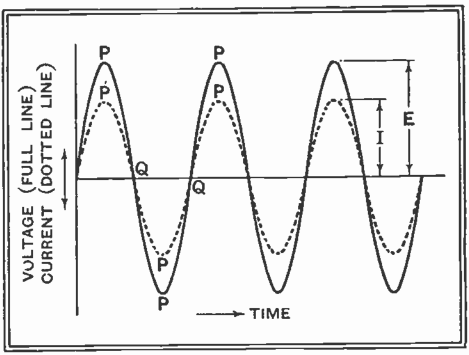


Fig. 29: Voltage and current relationships in the circuit of Fig. 28. The average power over one complete cycle is ½EI, or half the power developed when E and I both have maximum values

Alternating mains equivalent to 200-volt D.C. mains must therefore rise to a peak of $200\sqrt{2} = 282{\cdot}8$ volts. Such mains are described as having a *virtual* or R.M.S. (root-mean-square) voltage of 200.

If a fire of 40 ohms resistance is connected to such mains the R.M.S. current will be 200/40 = 5 amps., and the power consumed will be EI = 200 × 5 = 1,000 watts. Although the power is rapidly varying between a peak value of 2,000 watts and zero, the average power consumed and consequently the heat to which it gives rise will be exactly the same as if the same fire were connected to 200-volt D.C. mains.

Using in this way R.M.S. values for voltage and current we can forget entirely the rapid variations taking place, and *so long as our circuits are purely resistive* all calculations dealing with alternating current can be carried out according to the rules already discussed in connection with ordinary direct current.

33. Capacitance in A.C. Circuits

The behaviour of a condenser towards alternating current is best brought out by considering the effect of a number of successive charging currents in alternate directions. Imagine a circuit such as that of Fig. 30, consisting of a battery E, a condenser C, a meter M, and a rotary reversing-switch S. This latter is shown as four spring contact-arms, or *brushes*, numbered 1 to 4, pressing against the surface of a revolving drum or *commutator*. Except for the two segments, which are of metal, the commutator is supposed to be made of fibre, or other insulating material. In the position shown at *a*, the commutator serves to join, 1 to 2 and 3 to 4; when rotated through 90 degrees, as at *b*, it makes the connections 2 to 3 and 4 to 1. Turned again through a further 90 degrees, the connections at *a* are re-established. The resistance of the circuit is represented by *r*.

Tracing through the connections resulting from these

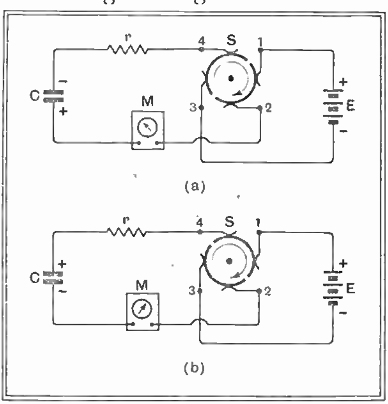


Fig. 30: Condenser and battery, with rotating reversing switch. The behaviour of this circuit leads directly to the properties of capacitance in an A.C. circuit

two positions of the commutator, it will be observed that in position *a* the upper plate, and in position *b* the lower plate, of the condenser is connected to the negative side of the battery.

Suppose that the circuit is first set up with the commutator as at *a*. At the exact moment of completing the circuit the condenser is as yet uncharged and there can therefore be no voltage across it; the whole of the E.M.F., E, is occupied in causing a charging current to flow against the resistance *r*. If *r* is 200 ohms and E is 100 volts the initial current must be 0·5 amp. This charging current immediately causes a voltage to be built up across C, and there is so much the less to drive the current through *r*. (Remember Kirchhoff's Law, by which the voltages across C and *r* when added at any instant must be equal—in this case—to 100.) The charging current therefore falls off, and, as the condenser nears its full charge, becomes very small indeed. So meter M shows a momentary deflection only. Reversing the connections by a quarter-turn of the commutator will connect the positive side of the battery to the negatively-charged side of the condenser; the total voltage acting to drive current round the circuit is 200—100 supplied by the battery and 100 by the charged condenser. So the current starts at 1 amp (see Fig. 31) and charges the condenser in the opposite polarity. M will record this by showing a large deflection in the opposite direction to the first.

If the commutator is turned slowly the meter will kick, first one way and then the other, every time the direction of connection is changed. By speeding up the rotation it will be found possible to make these alternations of direction so fast that the needle of the meter remains stationary in its central position through sheer inability to follow the successive kicks of current.

But if we replace this meter by another so designed that it deflects always in the same direction, no matter which way the current flows, the successive deflections produced by slow rotation of the commutator will simply fuse together as the speed of rotation is increased, the sluggishness of the meter preventing it from falling back to zero between successive rushes of current. We shall then have evidence of a current flowing, apparently continuously, in a circuit which is broken by the insulating material between the plates of the condenser. But, as the way in which the current has been built up clearly shows, electrons are flowing *in and out of* the condenser, and not *through* it in the ordinary sense of the word.

During each momentary burst of current the flow is greatest at the beginning and tails off towards the end, as the curve of Fig. 31 shows. The more rapid the rotation of the commutator, therefore, the greater is the proportion of the total time during which the current is high, and the greater, in consequence, is the average current read on the meter.

We conclude that the higher the frequency of the applied voltage, the greater will be the current in a circuit containing capacitance. It is easy to see, too, that, if the capacitance is greater, a greater number of electrons is required to perform the alternating charges and therefore the current is greater.

Instead of taking this rapidly reversing current from a

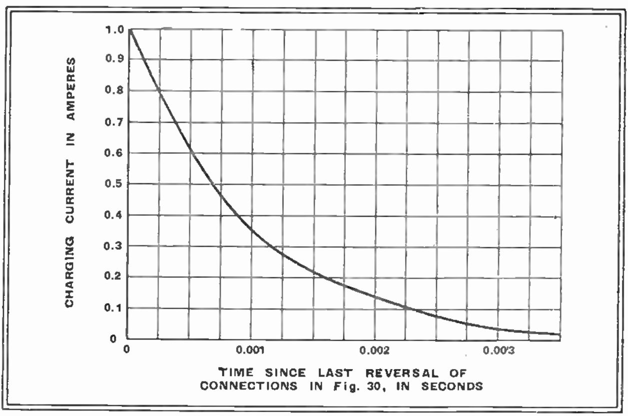


Fig. 31 : Showing the rapidity of the decay of the current after each reversal of direction

Calculated for $E = 100$ v., $C = 5\ \mu F$, $r = 200\ \Omega$

battery and a mechanical switch, it can be drawn from any normal source of alternating current, such as the electric light mains. If, as suggested in Fig. 32, a lamp (40-watt is recommended for the experiment) is connected to A.C. mains through a condenser of capitance some 2 μF or more, the lamp will light, and stay alight. But its brilliance will be below normal.

In the absence of the condenser, the alternating current drives electrons to and fro in the lamp filament ; with the condenser in circuit, the elastic opposition of the electrons in the dielectric restricts, to some small extent, the number of electrons that can so move at each change in direction of the voltage.

34. Reactance of a Condenser

As has just been indicated, the obstruction offered by a condenser to the flow of current depends upon its capacitance and upon the frequency of the current, becoming less as either of these rises. If an alternating potential of R.M.S. voltage E at a frequency f cycles per second is applied to a condenser of capacitance C farads the current flowing through it is $E \times 2\pi f C$ amperes R.M.S., where π is the ratio of the circumference of a circle to its diameter. The numerical value of this is 3·1416, or 22/7 approximately. A resistor to pass the same current would have a resistance of $1/2\pi f C$ ohms ; this figure is called the *reactance* of the condenser to currents of frequency f, and is expressed in ohms. In the case of the 2 μF condenser of Fig. 32, the reactance to 50-cycle current will be $1/2\pi f C = 1/2\pi.50.2.10^{-6} = 10^6/200\pi = 1590$ ohms.

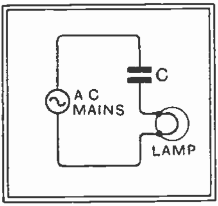


Fig. 32 : If a condenser of large capacitance (2 μF or more) is placed in series with a lamp lighted from A.C. mains, the lamp will light, though with less than normal brilliance

It is particularly to be noted that the electricity pushed into the condenser at one instant bounces out again the next ; the passage of an alternating current through a condenser *does not involve the expenditure of energy.* Such energy as is required to charge the condenser may be regarded as an extremely short-term loan, which is repaid in the next half-cycle, and so the net expenditure taken over any complete cycle is nil. If a resistance were used in place of C in Fig. 32 it would get hot, showing that this method of dimming the lamp diverts some of the unwanted energy to the resistance and there wastes it. Equal dimming by using a condenser wastes no power, as can be shown by the fact that C remains stone cold.

It is for this reason that its opposition to the current is called reactance, to distinguish it from resistance, the passage of current through which always involves the expenditure of energy.

35. Losses in Condensers

It is only in the ideal case, however, that the energy returned to the circuit on discharge is fully equal to that stored in charging the condenser, just as it is only a theoretically perfect spring that expands perfectly after compression. Imagine a "spring" made of copper wire, for example. So far from returning the loan with interest, as is usual in financial circles, a certain deduction is made on repayment. Since energy is lost when a current flows through such a condenser, it must possess resistance as well as reactance. This can be expressed, as in Fig. 33, by adding a resistance, either in series or in parallel, to the simple symbol for capacitance. The energy lost in such a composite circuit depends on the resistance alone, and

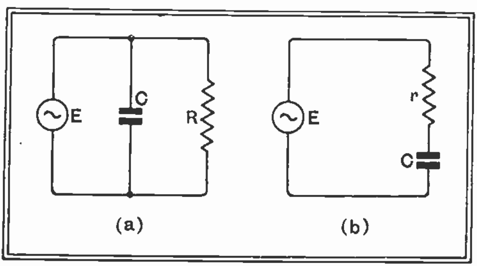


Fig. 33: A condenser which absorbs energy when a current flows through it can be represented as a perfect condenser with a resistance either in parallel (a) or in series (b). Watts lost are (a) E^2/R or (b) I^2r. R and r are related by the equation $\frac{1}{2\pi fC.R} = 2\pi fC.r$

can be calculated, in case *a*, by the formula $W = E^2/R$, where E is the voltage across both condenser and resistance, and in case *b*, by the formula $W = I^2r$, where I is the current flowing through the two in series.

Besides imperfections of dielectric, a further source of energy loss in a condenser is found in the resistance of the connecting wires and of the plates themselves. The condenser of Fig. 32, if used at a frequency of 1,500 kc/s, will have a reactance of $1/2\pi \cdot 1500 \cdot 10^3 \cdot 2 \cdot 10^{-6} = 1/6\pi = 0{\cdot}053$ ohm. Connecting wires and plates are evidently likely to have a resistance of at least this value, so that although they can be ignored at 50 cycles, where the reactance is 1,590 ohms, they may play a big part in the behaviour of the condenser at radio-frequencies.

36. Inductance in A.C. Circuits

It will be remembered (Sec. 19) that the characteristic of inductance is to delay the rise or fall of a current in a circuit, this being due to the formation of collapse of a magnetic field. If we imagine an inductance coil replacing the condenser of Fig. 30, making the circuit of Fig. 34, then on first com-

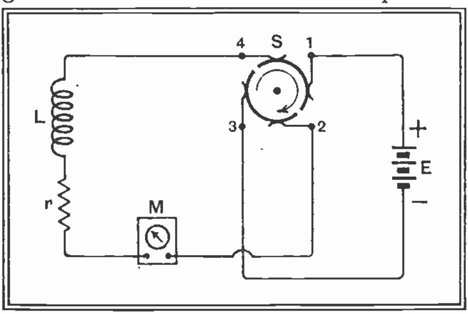


Fig. 34 : Inductance and resistance connected, through a reversing switch, to a battery

pleting the circuit the current will grow in the manner shown in the curve of Fig. 35. At sufficiently slow speeds of rotation of the commutator the total time taken by the growth of the current in alternate directions will be negligible compared with the time of steady flow, and the average current will be practically that which the resistance alone would take from the battery. At a higher speed, reversal might take place each time as soon as the current had risen to the value A of Fig. 35; the average current will now be smaller, but still considerable. By increasing the speed

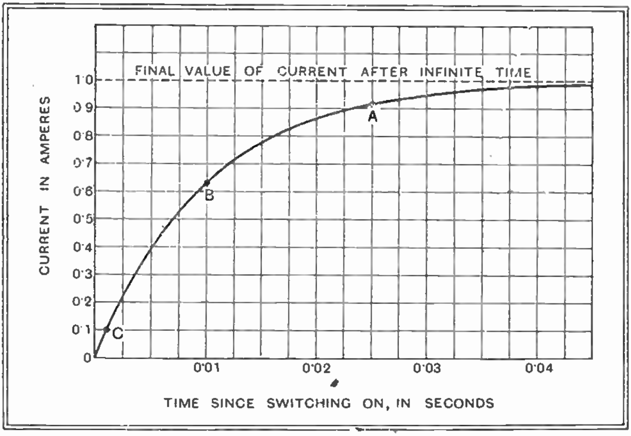


Fig. 35 : Showing slow rise of current in circuit of Fig. 34. Calculated for E = 100 v, L = 1·0 H, r = 100 Ω. By sufficiently rapid rotation of the switch, the current could be kept below A (1/40th sec.), below B (1/100th sec.), or even below C (1/1,000th sec.)

the reversal might be made so frequent as to prevent the current from ever exceeding B, or even C. It is clear that the greater the frequency of reversal the *less* will be the average current.

Compare this with the current through a condenser where, as Fig. 31 shows, sufficiently rapid alternation will prevent the current from *falling below* any chosen limit.

When an alternating voltage is applied to a coil the current that flows will be determined both by the frequency of the applied voltage and by the inductance of the coil, decreasing as either of these is raised. The resistance needed to take the same current, at a frequency f, as a coil of inductance L, is $2\pi f L$ ohms, where L is in henrys and f in cycles per second. This value is therefore the reactance of the coil to currents of frequency f.

So there are two kinds of reactance—capacitive and inductive—often denoted by X_C and X_L respectively. X is the general symbol for reactance.

As in the case of the condenser, no power is consumed by driving an alternating current against the opposition that this reactance represents, because the energy put into the magnetic field in building it up is restored to the circuit when it collapses. The resistance of the wire with which the coil is wound involves, of course, the usual consumption of energy, being I^2r, where I is the current flowing.

An inductance consists in most cases of a coil of wire. As a tuning coil it has been usual, till recently, to wind the coil on a tubular former of bakelite or cardboard ; some 100 turns of wire on a former of 1½in. diameter provide 170 μH or thereabouts for tuning over the medium wave band. A radio-frequency choke, of inductance perhaps 200,000 μH, will generally be wound of many turns of fine wire on a slotted former, though it may be a self-supporting coil of "wave-wound" type. Such a choke will offer a reactance of 1·26 megohms at f = 1,000 kc/s, while having a reactance of only 6,290 ohms at 5,000 cycles per second. Such a component is called a radio-frequency choke for the rather obvious reason that it opposes, or chokes back, the flow of currents of radio-frequency, while allowing those of audio-frequency a relatively unimpeded passage.

If it is necessary to offer considerable reactance to currents of quite low frequency, it is evident that a much higher inductance than this is necessary. To obtain high inductance without excessive resistance the coil is wound round a core of iron, or iron alloy, which offers a much easier passage than air to the lines of magnetic force, and so, by increasing the magnetic field, puts up the inductance which is a manifestation of that field.

37. Condensers in Parallel

The larger the capacitance of a condenser the greater the quantity of electricity required to charge it to a given voltage. Note the contrast : more ohms or more henrys mean *less* current, whereas more farads mean *more* current.

Take, for example, the circuit of Fig. 36, where C_1 = 1μF., C_2 = 2μF. The charging currents flowing

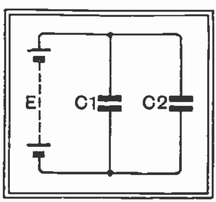


Fig. 36 : Condensers in parallel. Each condenser takes its own charging-current without respect to the other. Since condensers are rated by their charging-current, the capacitance equal to these two in parallel is given by $C = C_1 + C_2$

into the two condensers are quite independent ; when the current stops, each condenser has accepted the quantity of electricity necessary to charge it to the voltage of the battery. One microcoulomb into C_1, and two into C_2, for each volt of the battery ; total, 3 microcoulombs per volt, which, by the definition of capacitance, is the amount required to charge $3\mu F$. Evidently, the total capacitance of C_1 and C_2 taken together is the sum of the separate capacitances, and we may formulate the rule : If several condensers are connected in parallel, they make up a total capacitance equal to the sum of their individual capacitances.

$$C = C_1 + C_2 + C_3 + \ldots\ldots\ldots$$

Observe that the rule for condensers *in parallel* has the same form as that for resistances or reactances *in series*.

We can arrive at the same conclusion by simple algebraic reasoning based on the behaviour of the condensers to alternating current. Imagining the battery replaced by an A.C. generator of voltage E, then the currents through the condensers are respectively $E \times 2\pi f C_1$ and $E \times 2\pi f C_2$. The total current is thus $E \times 2\pi f (C_1 + C_2)$ which is equal to the current that would be taken by a single condenser of capacitance equal to the sum of the separate capacitances C_1 and C_2.

38. Condensers in Series

In Fig. 37 are shown two condensers connected in series. If X_1 and X_2 are respectively the reactances of C_1 and C_2, their combined reactance X is clearly $(X_1 + X_2)$, as in the case of resistances in series. By first writing down the equation "$X = X_1 + X_2$", and then replacing each "X" by its known value, of form $1/2\pi f C$, it is easy to see that $1/C = 1/C_1 + 1/C_2$.

That is, the sum of the reciprocals of the separate capacitances is equal to the reciprocal of the total capacitance. As in the case of resistances (Sec. 16) this equation can be put in a more manageable form—$C = C_1 C_2/(C_1 + C_2)$.

Observe that the rule for capacitances in series is identical in form with that for resistances or reactances in parallel.

From the way the rule was derived it is evidently not limited to two capacitances only, but applies equally to three, four, or more, all in series.

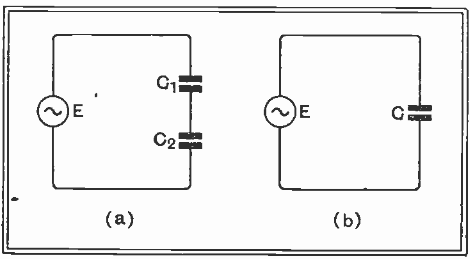
The rule implies that if two or more condensers are connected in series, the capacitance of the combination is always less than that of the smallest.

39. Inductances in Series or Parallel

The case of inductances in combination is quite straightforward. Several in series have a total inductance equal to the sum of their separate inductances, while if connected in parallel they follow the "reciprocal law", so that $1/L = 1/L_1 + 1/L_2 + 1/L_3 + \ldots\ldots$ In either case they combine just as do resistances or reactances. The reader can verify this for himself by adding reactances when they are in series, and adding currents when they are in parallel.

It should be noted that if L1 and L2 are placed in series their total inductance will only be $(L_1 + L_2)$ on the condition that the field of neither coil affects the other, If the coils interact so that their mutual inductance is M. the total inductance will be $(L_1 + L_2 + 2M)$ or $(L_1 + L_2 - 2M)$, "2M" being added or subtracted according to the direction of connection of the coils.

Fig. 37: If the one condenser C in *b* is to take the same current as the two in series at *a*, its capacitance will be given by $1/C = 1/C_1 + 1/C_2$



CHAPTER 5

A.C. CIRCUITS

40. Phase-Relations between Current and Voltage

WE have seen (Sec. 30), by means of a purely qualitative mental picture of the storage of energy in magnetic and electric fields, how an alternating current can pass through an inductance or a capacitance without dissipating energy. In order to make it possible to consider combinations of these with one another, or with resistance, we must make this picture much more precise, for we need to know the relationships between current and voltage that make this *wattless current* possible.

41. Resistance: E and I in Phase

In Fig. 38 the upper full line represents an alternating potential of R.M.S. value 1 volt. Rather more than one complete cycle is shown, and for convenience in reference each cycle is shown divided in the conventional way into 360 parts, corresponding to the 360 degrees of angle into which a circle is divided.

If this voltage is applied to a 2-ohm resistance the current will be $E/R = 0{\cdot}5$ amp. R.M.S. We have already seen that in a circuit consisting of pure resistance the current adapts itself instantaneously to changes in voltage; we may therefore apply Ohm's Law to each momentary voltage all through the cycle. By doing this we arrive at the dotted curve, which shows the current in the circuit at every instant. At the beginning of the cycle (at 0°) the voltage and current are both zero; at 90°, the end of the first quarter-cycle, both are at their maximum in a positive direction, dropping again to zero half-way through the cycle (at 180°), and rising again to a maximum in the negative direction.

Having thus drawn out voltage and current separately for each instant, we can calculate, by simple multiplication

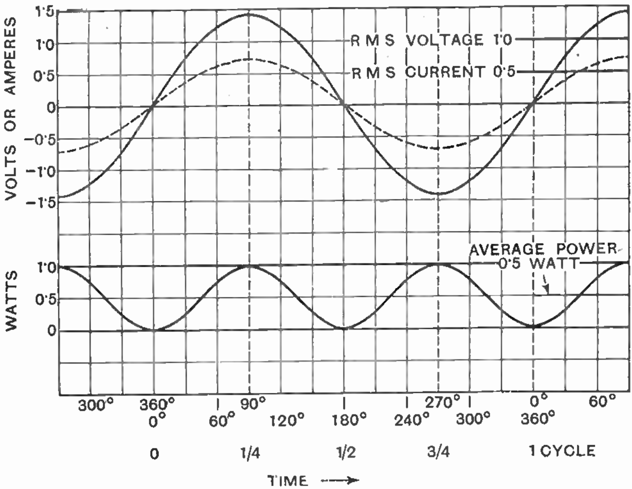


Fig. 38 : Relation of voltage, current, and power in a purely resistive circuit

of one by the other, the power being consumed. At 0°, for example, E and I are both zero ; so therefore is the power. At 30° E = 0·707, I = 0·353 ; hence the power EI, is 0·25 watt. Proceeding in this way for a number of points distributed over the first 180° of the cycle we find that the power rises to a maximum at 90°, and then falls again to zero, as the lower curve of Fig. 38 shows. In the next half-cycle, 180° to 360°, voltage and current are both negative ; their product is, therefore, still positive. A second rise and fall of wattage, exactly equal to that of the first half-cycle, will, therefore, occur.

In a resistive circuit, then, the power rises and falls once every half-cycle of the applied voltage. But it remains always positive, so that at every individual instant (except at 180° and 360°) power is being consumed in the circuit. R.M.S. voltage and current, and average power, are marked on the curves ; it will be seen that, as already explained, the calculation of average power from R.M.S. voltage and current, or from either of these and the value of resistance in the circuit, is worked out exactly as for direct current.

When current and voltage rise and fall exactly in step, as in the figure we have been discussing, the two are said to be *in phase*. It is evident that in any such case their product will remain positive at every instant. This relationship of current and voltage, therefore, cannot apply to wattless circuits (inductance or capacitance alone). In such circuits it is evident that the two must be out of step.

42. Capacitance : I Leading by 90°

In Fig. 39 is repeated the full-line voltage-curve of Fig. 38, but this time there is associated with it a current-curve displaced by 90°, or one-quarter of a cycle, towards the left. Since the diagram is read from left to right, this means that the current reaches its maximum a quarter of a cycle sooner than the voltage ; it is therefore said to

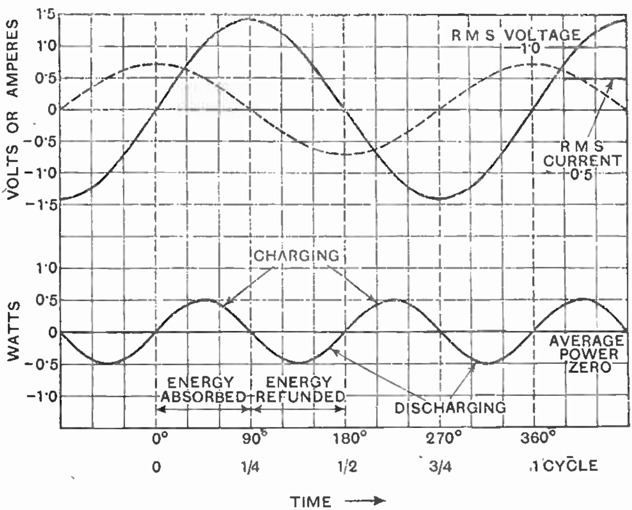


Fig. 39 : Relation of voltage, current and power in a purely capacitive circuit. Note that the average power is zero, and compare Figs. 38 and 40

lead the voltage by 90°, and is referred to as a *leading current*.

To calculate the power consumed with this new relationship between them we have, as before, to multiply corresponding pairs of values and plot the result. This leads to the lower full-line curve of this figure, from which it will be seen that the power is positive (i.e., absorbed) for the first quarter-cycle from 0° to 90°, is negative (i.e., refunded) for the next quarter-cycle from 90° to 180°, and so continues alternately positive and negative. This would correspond satisfactorily with the conditions known to hold when an alternating voltage is applied to a condenser *or* a coil, energy being alternately stored in, and returned from, the electric or magnetic field.

That the curves shown actually represent the case of the condenser can be seen if we remember that at every instant the voltage across it is that of the full-line curve. At moments of full charge the voltage across the condenser is at its maximum, but the current is zero, for it is just on the point of changing direction. These instants occur at 90° and 270° in Fig. 39. Immediately after each stop the current is positive if the voltage is running up from negative to positive (270° to 90°) and negative if the voltage is running from positive to negative (90° to 270°). The curves thus show in detail the way in which an alternating voltage drives a current through a condenser.

A common difficulty is "*How*, if the E.M.F. causes the current, can the current lead the E.M.F.? How does the current know what the E.M.F. that produces it is going to be?" This question evidently assumes that if, for example, an alternating voltage began at the moment specified as 360° in Fig. 39 then the current would start at 270°. Actually both would start together, and it would only be after a number of cycles that they would get into the relative phases shown in Fig. 39. The current, then, is not really determined in some occult way by a voltage that is still in the future, but by the present and past voltage. The statement that the current leads the voltage by quarter of a cycle means that after a sufficient number of cycles for both to get into their stride, a current maximum occurs quarter of a cycle earlier than a voltage maximum, and must not be taken to imply that that current maximum is a direct result of the voltage maximum that has not yet taken place.

43. Inductance : I Lagging by 90°

If we displace the current-curve by 90° to the right instead of to the left of that representing the voltage, we arrive at the diagram of Fig. 40. Here, again, the power is alternately positive and negative, making, as before, an average of zero power over the complete cycle. These curves show the relationship between voltage and current

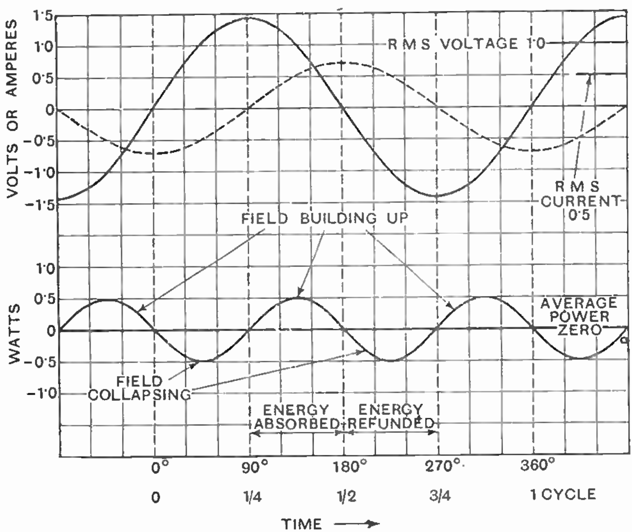


Fig. 40 : Relation of voltage, current and power in a purely inductive circuit. Note that the average power is zero, and compare Figs. 38 and 39

that is found when the circuit consists of pure (i.e., resistanceless) inductance. We have already seen that the need for building up the magnetic field round the coil slows the growth of the current, while its collapse tends to maintain the current for an instant after the voltage driving it is removed. Examination of the curves of Fig. 40 show that they fulfil just these conditions, the current rising and falling always later than the voltage. At 180°, for example, the current is flowing in a positive direction even though the voltage has dropped to zero. The current is in the direction in which the voltage was urging it a quarter of a cycle earlier.

As in the case of Fig. 39, current and voltage are *out of phase*, there being a *phase-difference* of 90° between them. This is the necessary condition for a wattless current. In the present case the current is known as a "lagging" current, for the reason that it reaches each maximum a quarter of a cycle after the voltage.

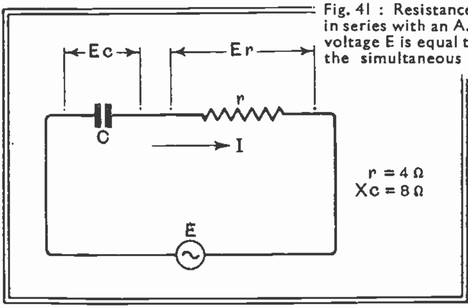


Fig. 41 : Resistance and capacitive reactance in series with an A.C. source. The generator-voltage E is equal to the total voltage due to the simultaneous presence of Ec and Er. See Fig. 42

44. Resistance and Capacitance in Series

Suppose a resistance and a capacitance are connected in series, and an alternating or high-frequency potential is applied across the whole, as in Fig. 41. It is evident that a current will flow, and that, since this current consists of the physical movement of electrons, it will be the same at all parts of the circuit at any one instant. A voltage, in phase with the current, will be developed across the resistance ; if the peak value of the current is 0·25 amp., as shown dotted in Fig. 42, the potential difference will rise to a maximum of 1 volt. This P.D. is shown as a full-line curve marked Er. Similarly, the current will develop a potential difference across the capacitance ; this, however, will be 90° out of phase with the current, as shown by the full-line curve Ec. Its maximum of 2 volts, therefore, does not coincide in time with the maximum of the voltage across the resistance.

The total voltage across the two circuit elements, which is, of course, equal to the voltage of the generator, must at every instant be equal to the sum of the two separate voltages (Sec. 12A), and can be found by adding the heights of the two curves, point by point over the cycle. (The term "adding", it is to be noted, may mean "subtracting" in the sense that + 1 v. and − ½ v. add up to + ½ v. by subtracting the negative half-volt from the positive volt.) The result of this addition is shown in the bottom curve of Fig. 42.

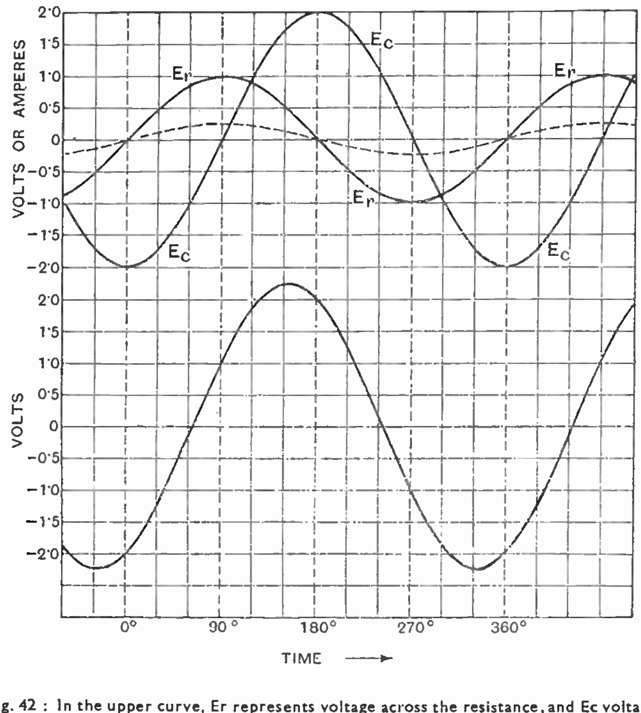


Fig. 42 : In the upper curve, Er represents voltage across the resistance, and Ec voltage across the condenser, of Fig. 41. The lower curve shows the resultant total voltage; that which the generator must have to drive 0·25 amp. through the circuit

It will be noticed that this total voltage has a phase between those of the two component voltages from which we have built it up; it is some 63° out of phase with the current. Further, the maximum voltage is not the sum of the two separate peak voltages, because these do not occur at the same instant, but, as it rises to 2·24 v., it is larger than either alone.

We are now in possession of the information that an alternating voltage of 2·24 v. drives a current of 0·25 amp. through a resistance of 4 ohms in series with a capacitance reactance of 8 ohms. The impedance of the circuit, which includes both resistance and reactance taken together is defined by clinging to the outward form of Ohm's Law and saying that the impedance Z shall be equal to the voltage divided by the current; i.e., $I = E/Z$ instead of $I = E/R$, as in the simple case of direct current. In our present case $Z = E/I = 2{\cdot}24/0{\cdot}25 = 8{\cdot}94\ \Omega$. The two components of the impedance, $8\ \Omega$ reactance plus $4\ \Omega$ resistance can obviously not be combined by simple addition to form this value of impedance, but it can be shown that a pure resistance r and a pure reactance X in series make up a total impedance Z which is given by $Z^2 = X^2 + r^2$. In our example, $Z^2 = 8^2 + 4^2 = 64 + 16 = 80$, whence $Z = 8.94\ \Omega$, as already found. An impedance worked out in this way can always be used in the "Ohm's Law" formula $I = E/Z$ to find the magnitude of the current that will flow on the application of a known voltage.

45. Resistance and Inductance in Series

This combination is almost identical with the last. In the circuit shown in Fig. 43 the generator E will drive some current I through the circuit. The voltage across r will be Ir, and this will be in phase with the current. Across the inductance the voltage will be IX, where X is the

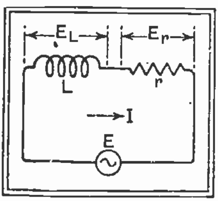


Fig. 43: Resistance and inductive reactance in series with a source of A.C. voltage. Compare with Fig. 41 and the curve of Fig. 42

reactance, $2\pi fL$, of the coil at the frequency of the generator. This voltage will be 90° out of phase with the current, and hence also 90° out of phase with the voltage Ir. These two voltages, Ir and IX, must together be equal to E, the voltage of the generator, but as their maxima do not occur at the same instant of time owing to their phase difference (see Fig. 42) their combined voltage is not the result of simply adding them together. Their phase-difference being exactly 90°, we can find E by combining them in the roundabout manner now beginning to become familiar: $E = \sqrt{(Ir)^2 + (IX)^2}$. This can also be written $E = I\sqrt{r^2 + X^2}$, showing that the impedance Z of this circuit is $\sqrt{r^2 + X^2}$.

Comparing this with the case of the condenser, we find that the same formula applies in both cases, since both are examples of the combination of voltages differing in phase by 90°. Problems involving an inductance in series with a resistance are therefore treated in exactly the same way as those depending on a resistance in series with a capacitance.

It is particularly to be noticed that inductance must always be associated with resistance in any practical case, even if the resistance is only the D.C. resistance of the wire with which the coil is wound. Although the two are in reality inextricably mixed up, it is satisfactory for purposes of calculation to regard any actual coil as a pure inductance in series with a pure resistance, as in the circuit of Fig. 43.

The resistance of the coil can also be represented in circuit diagrams by an equivalent parallel resistance R, such that $R = \dfrac{(2\pi fL)^2}{r}$ (compare Fig. 33). The usefulness of this will appear in Sec. 59.

46. Resistance and Reactance in Parallel

If a capacitance (or an inductance) is connected, in parallel with a resistance, across a source of alternating voltage, each branch will draw its own current independently of the other. These currents will be

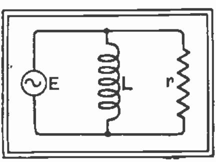


Fig. 44: Inductance and resistance in parallel. The combined impedance is given by
$1/Z = \sqrt{(1/r)^2 + (1/X)^2}$

E/r and E/X, where X is the reactance of the coil or condenser (Fig. 44). Since, like the voltages in Figs. 42 and 43, the two are not in phase, they cannot be added directly. That is, the magnitude of the total current is *not* equal to $E(1/r + 1/X)$. So long as the resistance is a pure resistance, and the reactance a pure reactance, so that the two currents are exactly 90° out of phase, the total current is given by $I = E\sqrt{(1/r)^2 + (1/X)^2}$. The impedance of r and X in parallel is, therefore, $\dfrac{1}{\sqrt{(1/r)^2 + (1/X)^2}}$

which may be simplified to $Z = \dfrac{rX}{\sqrt{r^2 + X^2}}$

We thus have :

Resistance and Reactance in Series : $Z = \sqrt{r^2 + X^2}$

Resistance and Reactance in Parallel : $Z = \dfrac{rX}{\sqrt{r^2 + X^2}}$

It is particularly to be noted that this addition of squares only applies to the simple case where the two currents or voltages involved are exactly 90° out of phase ; the further combination of one of these results with another reactance or resistance requires considerably more advanced methods than we propose to discuss here. The simple cases dealt with cover, fortunately, most ordinary wireless problems.

47. Power in A.C. Circuits

We have already seen that, in any alternating-current circuit, power is permanently consumed only when a current flows through a resistance. In this case voltage and current are in phase, and the power is equal to the product EI, both being expressed in R.M.S. units. In any purely reactive circuit current and voltage are 90° out of phase, and the power consumed, taken over a complete cycle, is zero. When both resistance and reactance are present together the phase difference lies between 90° and 0° as in Fig. 42, from which we conclude that the power consumed lies between zero and the product EI. It can be calculated by multiplying EI by a factor, always less than 1, that depends on the phase angle. But it is usually easier to find the current flowing through the circuit as a whole, and to multiply this by the voltage dropped across the resistive elements, ignoring entirely the voltage lost across the capacitance or inductance.

If, in Fig. 43, $X = 100\ \Omega$ and $r = 100\ \Omega$, the total impedance $Z = \sqrt{100^2 + 100^2} = 141 \cdot 4\ \Omega$. If $E = 200$ v., $I = E/Z = 1 \cdot 414$ amps. The voltage dropped across the coil is $IX = 141 \cdot 4$ v., but as it is known to be at 90° to the current, no power is consumed here. Across the resistance the voltage-drop is $Ir = 141 \cdot 4$ v., implying the consumption of $I \times Ir$ or $I^2r = 200$ watts. This is the sole consumption of power in the circuit. The same method of determining the power can be applied to any complex circuit in which we may happen to be interested.

In brief, power in an A.C. circuit can always be reckoned from the formula I^2r, but the alternative formulæ EI and E^2/r can only be used on the strict understanding that E stands for the voltage on the resistance alone. Further, the symbol r means resistance only, and does *not* mean total impedance.

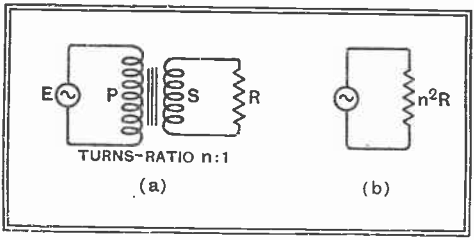


Fig. 45: Iron-cored transformer drawing current from the generator E and delivering it, at a voltage equal to E/n, to the load R. From the point of view of loading the generator, diagram b represents an equivalent circuit (assuming a perfect transformer)

48. Transformers

The mutual inductance between two circuits, discussed in Sec. 25, is widely used in all electrical work for transferring power from one circuit to another when it is desired that no direct metallic connection should exist between them. Suppose we have two coils, wound over the same iron core in such a way that as far as possible all the lines of force generated by passing a current through one coil will be led through the other. We then have a *transformer*, the symbol for which is seen in Fig. 45. The lines between the coils represent the iron core; an air-core transformer, consisting of two coils close together, would be indicated by the same symbol without these lines.

If alternating current is supplied to the primary winding P, the current through it is continuously rising and falling in alternate directions, with the result that P is surrounded by a continuously varying magnetic field. In the secondary winding S there is consequently set up a voltage, rising and falling in step with the changes in the field. With such *close coupling* between the coils that all the magnetic flux (lines of force) from P passes through S, the voltages across P and S will be equal if they have the same number of turns. If S has more or fewer turns, the voltage developed across it will stand to that across P in the ratio of the turns. There may be 2,000 turns in P and 40 in S; then if P is connected to 200-volt alternating mains S will deliver 4 volts A.C., and may be used to light the filaments of 4-volt valves. In practice it is not uncommon to have several secondary windings, delivering different voltages for different purposes, on the same iron core, all energized by a single primary.

In all calculations in which transformers enter, the relative voltages and currents in the two windings have to be computed on the basis of equal power, allowing, if necessary, for the small discrepancy due to losses in the transformer itself. If we have a two-to-one step-down transformer giving 100 volts at 1 amp. on the secondary side the power output is 100 watts. The useful power input to the primary is also 100 watts, so that from 200-volt mains the current taken would be ½ amp. The load on the secondary in this case is very clearly 100 ohms; the equivalent load on the primary, to take ½ amp. at 200 volts, would be 400 ohms. In general, if the turns-ratio of the transformer is n to 1, a resistance R_s connected across the secondary winding has the same effect on the primary as if there were connected across it a resistance R_p, equal to n^2R_s. The turns-ratio is therefore given by $n^2=\frac{R_p}{R_s}$, or $n=\sqrt{\frac{R_p}{R_s}}$.

We shall see in Chapters 10 and 11 that this simple conclusion has a very important application to the design of amplifying stages in a wireless set.

Unless it is desired to insulate the secondary winding from the primary, it is not necessary to have two separate windings. The winding having the smaller number of turns can be abolished, and the connections tapped across the same number of turns forming part of the other winding. This device is called an *auto-transformer*.

CHAPTER 6

THE TUNED CIRCUIT

49. Inductance and Capacitance in Series

THE use of inductance and capacitance in combination is one of the outstanding features of any wireless circuit ; it will therefore repay us to make a fairly close study of their behaviour. Fig. 46 *a* shows a coil and a condenser connected in series across an A.C. generator of voltage E. As in the case of Fig. 42 (C and *r* in series) we will begin with the current and work backwards to find the voltage necessary to drive it.

The dotted curve of Fig. 47 represents, by the usual sine-curve, the current flowing ; we have allotted to it a value of 0·25 amp. peak. Taking the reactance of C as 4 Ω, the voltage across it will be 1 v. peak, displaced in phase by 90° from the current. The rise and fall of this voltage with time is given by the full-line curve E_C. As required for a condenser, the current reaches its maximum a quarter of a cycle before the voltage.

The voltage across the coil, of reactance 8 Ω, will be 2 v. peak, and its phase will be such that the current reaches each maximum a quarter of a cycle after the voltage. The full-line curve E_L fulfils these conditions.

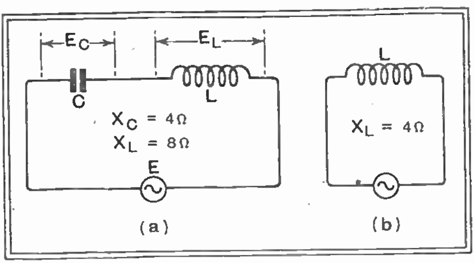


Fig. 46 : Diagram *a* represents inductance and capacitance in series across an A.C. source, while diagram *b* shows the equivalent circuit for the values given. See Fig. 47

It will be seen at once that the two voltages E_L and E_C are out of phase by 180°, which means that at every instant they are in opposition. If we find the sum of the two by adding the heights of the curves point by point and plotting the resulting figures we obtain for E (the generator voltage necessary to drive the assumed quarter-ampere through the circuit) the curve at the bottom of the diagram.

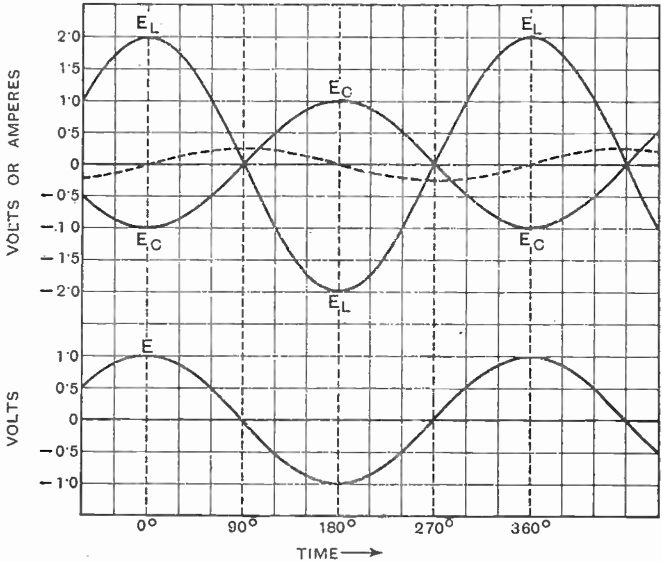


Fig. 47 : E_C and E_L represent the voltages across C and L of Fig. 46 when the current shown by the dotted curve is flowing. These voltages are at every instant in opposition and together make up to the voltage E

In obtaining this curve it was necessary to perform a subtraction at each point, since the two component voltages are at every instant in opposition. It is, therefore, scarcely surprising to find that the voltage required for the generator has the phase of the larger of the two voltages and is equal in magnitude to the difference of the two. The peak value of E is 1 volt, and its phase with respect to the current is that of the voltage across the inductance.

The same voltage and current relations that we find for the complete circuit could therefore equally well have been produced by applying 1 volt to a coil of reactance 4 ohms, as in Fig. 46 *b*. The capacitive reactance X_C of 4 Ω has exactly nullified 4 of the original 8 ohms of the inductive reactance X_L, leaving 4 Ω of inductive reactance still effective. We therefore conclude that the total reactance of a series combination of L and C is given by: $X = X_L - X_C$.

If we had made $X_L = 4\ \Omega$ and $X_C = 8\ \Omega$ in the original example, we should have found the circuit equivalent to a condenser of reactance 4 ohms. Applying the same rule, the total reactance would now be $(X_L - X_C) = (4 - 8) = -4\ \Omega$. As a physical entity, a negative reactance is meaningless, but the statement of the total reactance in these terms is, nevertheless, accepted as correct, the minus sign being conventionally taken to indicate that the combined reactance is capacitive.

50. L, C and *r* all in Series

Since the combination of a coil and a condenser in series is always equivalent either to a coil alone or to a condenser alone, it follows that the current through such a combination will always be 90° out of phase with the voltage across it. We can therefore combine the whole with a resistance in the same manner as any other reactance. To find the total impedance of the circuit of Fig. 48 *a*, for example, we have first to find the reactance X equivalent to X_L and X_C taken together; $X = X_L - X_C$. To bring in the resistance we use the formula $Z = \sqrt{X^2 + r^2} = \sqrt{(X_L - X_C)^2 + r^2}$. There is no more complication here than in combining a resistance with a simple reactance.

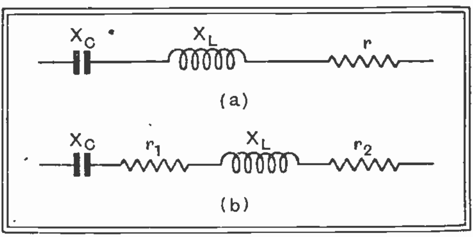


Fig. 48: Capacitance, inductance and resistance in series. For *a*, $Z^2 = (X_L - X_C)^2 + r^2$. For *b*, $Z^2 = (X_L - X_C)^2 + (r_1 + r_2)^2$

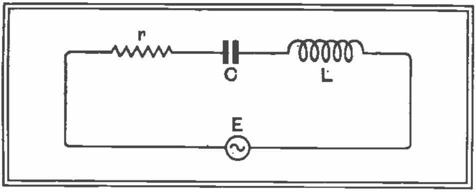


Fig. 49 : Series-tuned circuit : if L=200μH, C = 200$\mu\mu$F, r = 10 ohms, the magnification will be 100 at resonance. (See Fig. 50)

Faced with a circuit like that of Fig. 48 *b*, we might feel inclined to begin by combining r_1 with X_C and r_2 with X_L, afterwards combining the two results. But a little consideration will show that neither of these pairs would be either a pure resistance or a pure reactance, so that we should have no immediate knowledge of the relative phases of the voltages across them. The final stage of the process would, therefore, be outside the range of the methods we have discussed. We get round the difficulty by first finding the total reactance of the circuit by adding X_L, and X_C, then finding the total resistance by adding r_1 and r_2, and finally working out the impedance as for any other simple combination of reactance and resistance. The fact that neither the two reactances nor the two resistances are neighbours in the circuit does not have to be taken into consideration, since the same current flows through all in series.

51. The Series-Tuned Circuit

We have already seen (Sec. 33) that the reactance of a condenser falls and that of an inductance rises (Sec. 36) as the frequency of the current supplied to them is increased. It is therefore going to be interesting to study the behaviour of a circuit such as that of Fig. 49 over a range of frequencies. For the values given for the diagram, which are reasonably representative of practical broadcast reception, the reactances of coil and condenser for all frequencies up to 1,800 kilocycles per second are plotted as curves in Fig. 50. The most striking feature of this diagram is that at one particular frequency, about 800 kc/s, the coil and the condenser have equal reactances, each amounting then to about 1,000 ohms.

At this frequency the total reactance, being the difference of the two separate reactances, is zero. Alternatively expressed, the voltage developed across the one is equal to the voltage across the other; and since they are, as always, in opposition, the two voltages cancel out exactly. The circuit of Fig. 49 would, therefore, be unaltered, so far as concerns its behaviour *as a whole* to a voltage of this particular frequency, by the complete removal from it of both L and C. This, leaving only *r*, would result in the flow of a current equal to E/*r*.

Let us assume a voltage not unlikely in broadcast reception, and see what happens when E = 5 millivolts. The current at 800 kc/s will then be 5/10 = 0·5 milliamp., and this current will flow, not through *r* only, but through L and C as well. Each of these has a reactance of 1,000 ohms at this frequency; the potential across each of them will therefore be 0·5 × 1,000 = 500 mV., which is just one hundred times the voltage E of the generator to which the flow of current is due.

That so small a voltage should give rise to two such large voltages elsewhere in the circuit is one of the queer

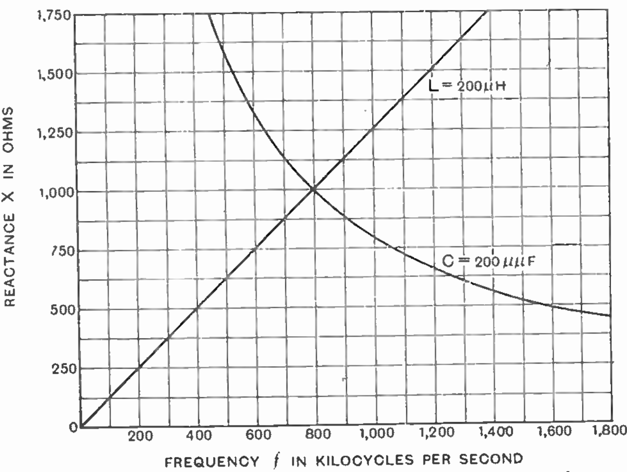


Fig. 50: Reactances of the coil and condenser of Fig. 49 plotted against frequency. Note that 800 kc/s, where the curves intersect, is the frequency of resonance

paradoxes of alternating currents that make wireless possible. If the foregoing paragraphs have not made clear the possibility of the apparent absurdity, the curves of Fig. 47, modified to make the two voltages equal, will give the complete picture of the large individual voltages in opposite phase.

52. Magnification

In the particular case we have discussed, the voltage across the coil (or across the condenser) is one hundred times that of the generator. This ratio is called the *magnification* of the circuit, and is generally denoted by the letter Q.

We have just worked out the Q for a particular circuit; now let us try to obtain a formula for the Q of any circuit. It is equal to the voltage across the coil divided by that from the generator, which is the same as the voltage across r. If I is the current flowing through both, then the voltage across L is $2\pi f LI$, or X_L I; and that across r is Ir. So $Q = 2\pi f L/r$ or X_L/r. At any given frequency, magnification depends solely on L/r, the ratio of the inductance of the coil to the resistance of the circuit.

If r is made very small, the current round the circuit for the frequency for which the reactances of L and C are equal will be correspondingly large. In the theoretical case of zero resistance, the circuit would provide, at that one frequency, a complete short-circuit to the generator. Huge currents would flow, and the voltages on C and L would in consequence be enormous.

To obtain high magnification of a received signal (for which the generator of Fig. 49 stands), it is thus desirable to keep the resistance of the circuit as low as possible.

53. Resonance Curves

To voltages of frequencies other than that for which coil and condenser have equal reactance, the impedance of the circuit as a whole is not equal to r alone, but is increased by the net reactance. At 1,250 kc/s, for example, Fig. 50 shows that the individual reactances are 1,570 and 636 ohms respectively, leaving a total reactance of 934 ohms. Compared with this, the resistance is negligible, so that the current, for the same driving voltage of 5 mV., will be 5/934 mA, or, roughly, 5 microamps. This is approximately one hundredth of the current at 800 kc/s.

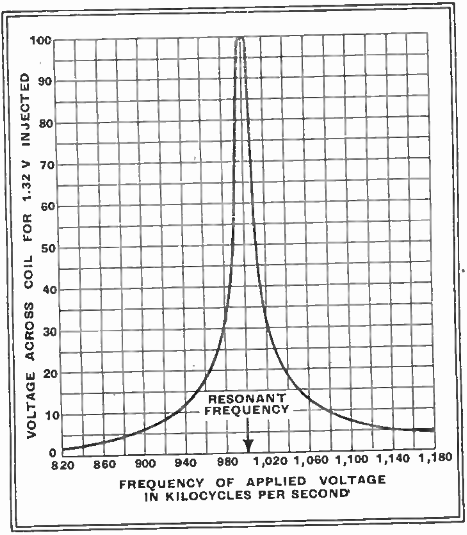


Fig. 51: Voltage plotted against frequency for a series circuit in which $L = 180\mu H$, $C = 141\mu\mu F$, $r = 15$ ohms, $V = 1{\cdot}32$ volts

By extending this calculation to a number of different frequencies we could plot the current in the circuit, or the voltage developed across the coil, against frequency. The curve so obtained is called a *resonance-curve*; one is shown in Fig. 51. The vertical scale shows the voltage developed across the coil for an injected voltage of 1·32 volts; at 1,000 kc/s, the frequency at which $X_L = X_C$, the voltage across the coil rises to 100 volts, from which we conclude that $Q = 100/1{\cdot}32 = 75$. Without going into details, a glance at the shape of the curve is enough to show that the response of the circuit is enormously greater to voltages at 1,000 kc/s than to voltages of substantially different frequencies; the circuit is said to be *tuned* to, or to *resonate* to, 1,000 kc/s.

The principle on which a receiver is tuned is now beginning to be evident; by adjusting the values of L or C in a circuit such as that under discussion it can be made to resonate to any desired frequency. Any signal-voltages received from the aerial at that frequency will receive preferential amplification ; and the desired transmitter, distinguished from the rest by the frequency of the wave that it emits, will be heard to the comparative exclusion of the others.

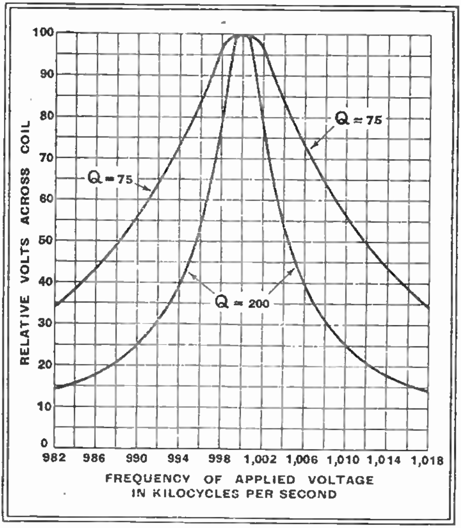


Fig. 52 : Resonance curves of two tuned circuits of different magnifications (Q) at 1,000 kc/s. The greater selectivity of the circuit of higher magnification is very apparent. Note that E (Fig.49) is 1·32 volts for coil Q = 75, but only 0·5 v. for coil Q = 200

54. Selectivity

We have said "comparative exclusion" because it is found that the *selectivity* of a single tuned circuit is seldom enough to provide sufficient separation between stations, so that two, three, or even more are used, all being tuned together by a single knob. The increase of selectivity obtained by multiplying circuits is very marked indeed ; with a single circuit of the constants of Fig. 51 a station is reduced to one-twentieth of its possible strength by tuning away from it by 120 kc/s (5 v. response on Fig. 51 at $f = 880$ or 1,140 kc/s). Adding a second tuned circuit to select from the signals passed by the first leaves only one-twentieth of this twentieth—i.e., one four-hundredth. A third circuit leaves one-twentieth of this again—that is, one eight-thousandth. This last figure represents a set of about the minimum selectivity acceptable for general reception ; it follows that a receiver requires a minimum of three tuned circuits except in cases where means are provided for increasing the sharpness of tuning beyond that given by the unaided circuit.

The sharpness with which a circuit tunes depends entirely upon its magnification, as comparisons of the two curves of Fig. 52 will show. These are plotted to the same maximum height, thereby helping comparisons of selectivity while obscuring the fact that a circuit of $Q = 200$ gives a louder signal (more volts at resonance) than one for which $Q = 75$. In Fig. 53 the curves are redrawn to show the relative response of the two circuits to the same applied voltage; the more selective circuit is also, as we have seen, the more efficient.

55. Resonant Frequency

At the frequency of resonance the reactance of the coil equals that of the condenser; consequently we know that for that particular frequency $2\pi f L = 1/2\pi f C$. By a little rearrangement of this equation, we get the important relationship $f = 1/2\pi\sqrt{LC}$, the resonant frequency f being in cycles per second, while L and C are in henrys and farads respectively. This formula

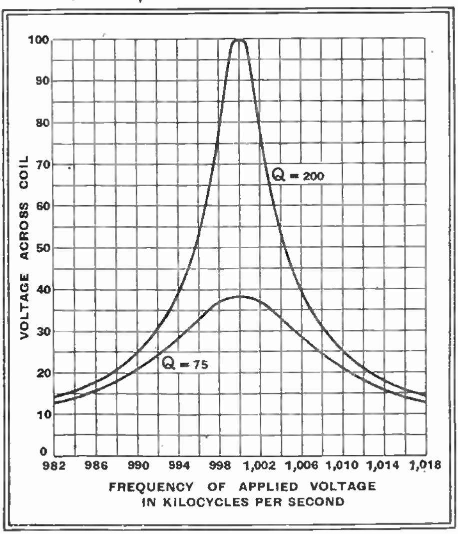


Fig. 53: Voltages on two different coils for the same injected voltage ($E = 0{\cdot}5$ v.). Note greater response, as well as higher selectivity (Fig. 52), of coil of higher Q

allows us to predict the frequency to which any chosen combination of inductance and capacitance will tune.

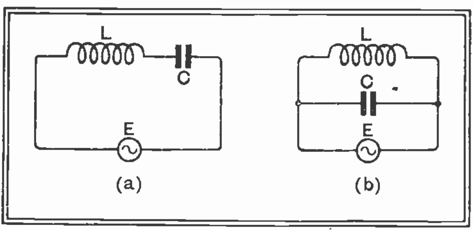
If we prefer our answer in terms of wavelength, we can replace f in the formula by its equivalent $\frac{3 \times 10^8}{\lambda}$, where λ is the wavelength in metres. This leads to the well-known formula $\lambda = 1{,}885\sqrt{LC}$, where the figure 1,885 includes all numerical constants, and is made a convenient number by taking L in *micro*henrys and C in *micro*farads.

It will be noticed that if a coil is tuned by a variable condenser (the customary method) it is necessary to quadruple the capacitance in order to double the wave-length or halve the frequency. This is so because wave-length is proportional to the square root of the capacitance.

The average tuning condenser has a maximum capacitance of about 530 $\mu\mu$F, while the minimum capacitance, dependent more on the coil and the valves connected to it than upon the condenser, is generally about 70 $\mu\mu$F in a modern set. This gives a ratio of maximum to minimum capacitance of $530/70 = 7{\cdot}57$. The ratio of maximum to minimum frequency is the square root of this, namely, 2·75. Any band of frequencies with this range of maximum to minimum can be covered with one swing of the condenser, the exact values of the frequencies reached being dependent on the inductance chosen for the coil.

Suppose we wished to tune from 1,500 kc/s to 1,500/2·75 or 545 kc/s, corresponding to the range of wavelengths 200 to 550 metres. For the highest frequency or lowest wavelength the capacitance will have its minimum value of 70 $\mu\mu$F; by putting the appropriate values in either

Fig. 54: Series- and parallel-tuned circuits compared. In the series circuit *a* the current through L and C is the same; in *b* the voltage across L and C is the same



the formula for f or that for λ we find that L must be made 161 μH.* It should be evident that if we calculate the value of L necessary to give 545 kc/s (550 metres) with a capacitance of 530 $\mu\mu$F, the same value will again be found.

By using instead a small inductance suitable for the short waves (0·402 μH.) we could cover the range 10 to 27·5 metres (30,000 to 10,900 kc/s, or 30 to 10·9 *mega*-cycles per second), while the choice of 2120 μH, (or 2·12 *milli*henrys) would enable us to tune from 728 to 2,000 metres.

Observe how convenience is served, large and clumsy numbers dodged, and errors in the placing of a decimal point made less likely by suitable choice of units, replacing "kilo-" by "mega-", or "micro-" by "milli-" whenever the figures suggest it. The preceding paragraph, rewritten in cycles and henrys, would be almost impossible to read.

56. The Parallel-Tuned Circuit

The series-tuned circuit rather obviously derives its name from the fact that the voltage driving the current is in series with both coil and condenser, as in Fig. 54 *a*. In its very similar counterpart, the parallel-tuned circuit, the voltage is considered to be applied in parallel with both coil and condenser, as in Fig. 54*b*. The change in circuit from one to the other results in a kind of interchange in the functions of current and voltage.

* Worked out thus. $f = 1/2\pi\sqrt{LC}$, so that $(2\pi f)^2 C = 1/L$, or $L = 1/(2\pi f)^2 C$. Now putting in values:

$$\begin{aligned} L &= 1/(2\pi \times 1500 \times 10^3)^2 \times 70 \times 10^{-12} \\ &= 1/88{\cdot}9 \times 10^{12} \times 70 \times 10^{-12} \\ &= 1/6210 = 0{\cdot}000161 \text{ henrys} \\ &= 161 \text{ microhenrys.} \end{aligned}$$

Starting from the formula $\lambda = 1885\sqrt{LC}$, we get $(\lambda/1885)^2 = LC$, or $L = \frac{1}{C}\left(\frac{\lambda}{1{,}885}\right)^2$. Putting in values,

$$\begin{aligned} L &= \frac{1}{70 \times 10^{-6}}\left(\frac{200}{1{,}885}\right)^2 \\ &= 0{\cdot}01125/(70 \times 10^{-6}) \\ &= 11{,}250/70 \\ &= 161 \text{ microhenrys, as before.} \end{aligned}$$

In *a*, the current is necessarily the same at all parts of the circuit; we elucidated its behaviour by considering the voltages that this current would produce across the various components and added them up to find E, the driving voltage. In *b*, the position is reversed; here we have the same voltage applied to both the inductive and the capacitive branches, and we have to find the separate currents in the two and add them to find the total current.

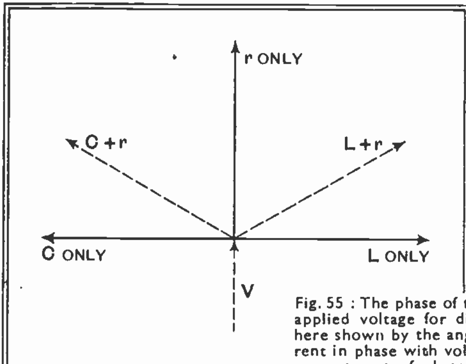


Fig. 55 : The phase of the current relative to the applied voltage for different types of circuit is here shown by the angle of a line : r gives current in phase with voltage V, while C and L give currents out of phase by 90° in opposite directions. Combinations of C and *r* or L and *r*, give intermediate phases. Note that the currents with C only and L only tend to cancel out in a parallel circuit

In the absence of resistance, the current in the L-branch will be determined by the reactance $2\pi f L$ of the coil; it will be $E/2\pi f L$. In the C-branch, it will similarly be $E/(1/2\pi f C) = E.2\pi f C$. We know already that these two currents will be exactly out of phase with one another, as were the voltages in Fig. 47. The net current taken from the generator will, therefore, be the simple difference of the two individual currents.

The currents become equal, and their difference consequently zero, at the frequency of resonance. As in the series circuit, this occurs when $2\pi f L$ equals $1/2\pi f C$, so that once again the frequency of resonance is given by $f = 1/2\pi\sqrt{LC}$. A coil and condenser thus tune to the same frequency irrespective of whether they are arranged in series or parallel with the source of voltage that drives the current.

A parallel circuit has a resonance curve in all respects similar to that of the series circuit already discussed, the sharpness of tuning being determined, as before, by the magnification, $2\pi f L/r$.

57. Series and Parallel Circuits Compared

In the series circuit, the current flowing produced across L and C two equal voltages, which, although they might individually be quite large, cancelled one another out. Taken together, L and C formed a part of a circuit across which no voltage was developed however large the current flowing; their joint impedance, therefore, was zero. If it were not for the presence of resistance, the series circuit would act as a short-circuit to currents of the frequency to which it is tuned; it is therefore often known as an "acceptor" circuit. Conditions are similar in the parallel resonant circuit. Here the voltage E produces through L and C two equal currents, which, although they may individually be quite large, cancel one another out. Taken together, L and C give a circuit through which no current flows however large the voltage applied; their joint impedance, therefore, is infinitely large.

The parallel circuit thus acts as a perfect barrier to the passage of currents of the frequency to which it is tuned; it is therefore often known as a "rejector" circuit.

58. The Effect of Resistance

But it will be clear that two conditions are necessary for this rejector action to be perfect. Firstly, the currents through the two branches must be equal, which can only happen when $X_L = X_C$; in other words, at the exact frequency of resonance. At other frequencies there will be more current through one branch than through the other.

The second condition for complete cancellation of the two currents is that they shall be out of phase by exactly 180 degrees. We have already seen that in a mixed circuit, containing both L and r or C and r the phase of the current lies between those appropriate to the individual circuit-elements in the way summarized in Fig. 55 and shown for one particular case in the curves of Fig. 42. It follows that if resistance is present in either the inductive or the capacitive branch of a parallel-tuned circuit, as in Fig. 56*a*, the two currents are less than 180 degrees out of phase, and so can never exactly cancel one another.

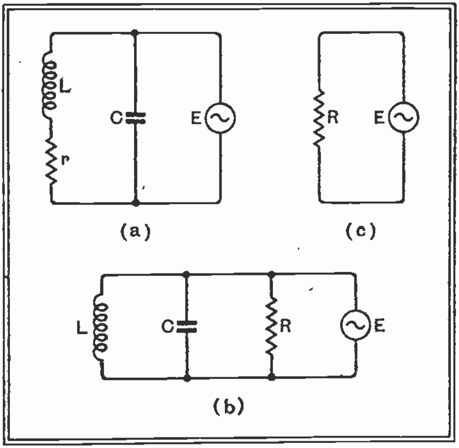


Fig. 56 a : Parallel-tuned circuit with resistance in coil. This is equivalent to *b*, which at the frequency of resonance can be simplified to c

Even at resonance, therefore, there is a small residual current, with the result that the tuned circuit no longer presents a *complete* barrier to the passage of currents of the frequency to which it is tuned. Further, it will be clear that the larger the resistance *r* of Fig. 56*a*, the more the phase of the current passing through that branch of the circuit will depart from that proper to a purely inductive circuit, and so the larger will be the uncancelled residue of the capacitive current. Put briefly, a larger *r* leads to a larger current through the circuit as a whole, and hence to a decrease in the total impedance of the circuit.

59. Dynamic Resistance

Sections 35 and 45 show that a real condenser or coil (which always has *some* resistance) can be represented in a circuit diagram by a pure capacitance or inductance in series *or* in parallel with a pure resistance. It is possible then, to redraw Fig. 56*a* as Fig. 56*b*, in which $R = (2\pi f L)^2/r$. We now have a resistance, R, in parallel with a *perfect* rejector circuit L and C, which passes no current at the resonant frequency, and can therefore be removed entirely, giving Fig. 56*c*. It is therefore permissible to replace the whole of a tuned circuit by a pure resistance R, it being strictly understood that this simplification is only allowable as long as we restrict ourselves to considering the behaviour of the circuit towards currents of the exact frequency to which it is tuned.

This resistance R, as we have seen, is infinitely large when r, the true resistance of the circuit, is zero, but decreases as r is increased. Since real, physical resistances do not behave in this topsy-turvy way, we have to distinguish R from an ordinary resistance by coining a special name for it; it is generally referred to as the *dynamic resistance* of the tuned circuit. As at the resonant frequency $2\pi f L = 1/2\pi f C$, it is possible to simplify the formula $R = (2\pi f L)^2/r$ to $R = L/Cr$. Actually this is not perfectly correct, because the conversion of Fig. 56*a* to *b* involved a change in the value of L and hence in the formula for resonant frequency; but except for very flatly tuned circuits the error is generally negligible. Thus a tuned circuit consisting of an inductance of 160 μH, tuned with a capacitance of 200 $\mu\mu$F, and with a high-frequency resistance r of 7 ohms has a dynamic resistance R = $(160 \times 10^{-6})/(200 \times 10^{-12} \times 7) = (160 \times 10^{6})/1{,}400 = 114{,}000$ ohms. It is evident that if r had been $3\frac{1}{2}$ or 14 ohms R would have come out at 228,000 or 57,000 ohms respectively, so that halving or doubling the resistance r doubles or halves the dynamic resistance.

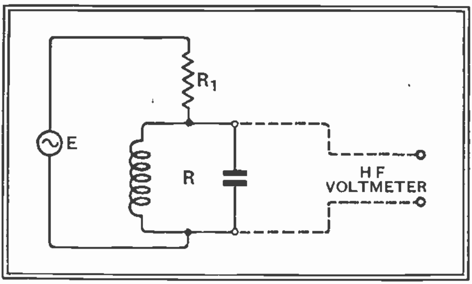


Fig. 57: Simple method of determining dynamic resistance for a parallel-tuned circuit. If R_1 is adjusted until the R.F. voltmeter reads E/2 volts, we know that $R_1 = R$

The relation between r and R is such that the specification of either, in conjunction with the values of L and C, completely determines the behaviour of the circuit at resonance.

If R is known, r can be found from the relation $r = L/CR$. Low values of parallel, or high values of series resistance *damp* the circuit, resulting in flat tuning.

60. Measuring R and r

In the sense that it cannot be measured by ordinary direct-current methods—by finding what current passes through it on connecting across it a 2-volt cell, for example —it is fair to describe R as a fictitious resistance. Yet it can quite readily be measured by such means as those outlined in Fig. 57, using for the measurement currents of the frequency to which the circuit is tuned. In spite of the inevitability of resistance in the windings of a coil, r is fictitious to just the same extent as R, for a true value of r cannot be obtained by any direct-current method. Indeed, it may often happen that a change in a coil that will reduce the resistance to direct current—by rewinding it with a thicker wire, for example—has the effect of increasing the high-frequency resistance instead of diminishing it. A true value for r can only be found by making the measurement at high frequency, using some such method as that outlined in Fig. 58.

It is possible to calculate the resistance offered to high-frequency currents by the wire with which a coil is wound. This value is always considerably higher than the plain resistance of the wire to ordinary direct current. From our present point of view the reasons for this particular discrepancy are not of much importance; we shall visualize them well enough by remembering that each turn of the coil lies in the magnetic field of the other turns, which has the result that there are set up stray currents in addition to the main current, thereby increasing the losses due to the resistance of the wire. Even in a straight wire the resistance at high frequency is greater than for steady currents, the magnetic field setting up the stray currents responsible for this being derived from the main current in the wire itself.

61. Dielectric Losses

By making a measurement of the high-frequency resistance of a tuned circuit on the lines indicated in Fig. 58 we always find a value for r which is very appreciably higher than that found by calculation. This indicates that there are sources of resistance other than the wire with which the coil is wound. Investigation shows that this additional resistance, which may even be greater than that of the coil, is due to imperfections in the dielectric materials associated with the tuned circuit.

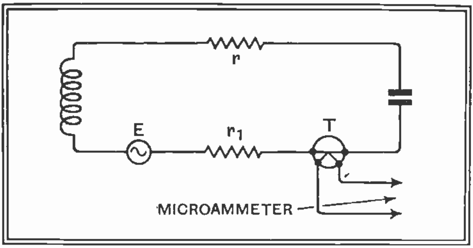


Fig. 58: Simple method of determining series resistance r. If the current is noted when $r_1 = 0$, and then reduced to half this value by introducing and adjusting r_1, the necessary value of r_1 is r. The resistance of the H.F. milliammeter (thermo-junction T) is included in r, and must be allowed for

The plates of the tuning condenser, for example, have to be supported in some way ; even if the dielectric between the plates is mainly air there is some capacitance between neighbouring portions of the two sets of plates for which the insulating support provides the dielectric. Valve-holders, valve-bases, or terminal blocks, connected across the tuned circuit also introduce capacitance, the dielectric again being the insulating material on which the metal parts are mounted.

All these dielectrics are imperfect in the sense that they are not " perfect springs ". In other words, in the rapid to-and-fro movement of electrons set up in them by the high-frequency voltage across the tuned circuit a certain amount of energy is absorbed and dissipated as heat. We have seen that the absorption of energy is an inseparable characteristic of resistance ; a circuit containing such sources of energy-loss as these is therefore found to have a high-frequency resistance r higher than that calculated for the coil and other metallic paths alone. The total is referred to as the "equivalent series resistance" of the circuit, the value of r so described being that which in conjunction with a perfectly loss-free condenser and a resistanceless coil would give a tuned circuit identical with the actual one at the frequency for which the measurements of resistance were made.

Physically, these dielectric losses behave as though they were a resistance in parallel with the circuit, as in Fig. 56*b*; but just as we converted a series resistance r into an equivalent parallel resistance, using the formula $R = L/Cr$, so we can convert back using $r = L/CR$. For example, suppose that a particular valve-holder is equivalent to 0·45 MΩ parallel resistance at 250 metres. If $L = 100\mu H$, then $C = 176\ \mu\mu F$, and the value of r added to the circuit by connecting the valve-holder across it is $100/(176 \times 0{\cdot}45) = 1{\cdot}26\ \Omega$. But if the inductance of the coil were 200 μH, the added series resistance equivalent to the valve-holder would be 5·04 Ω, four times the preceding value. (L doubled implies also C halved.) As the true series resistance of the 200 μH coil (i.e., the resistance actually due to the winding itself) will be approximately double that of the 100 μH coil at the same frequency, it follows that the damping effect due to the valve-holder will be twice as great when the larger coil is in use.

A true series loss, such as a high-resistance connection in a switch or at a soldered joint, will add the same series resistance irrespective of the inductance of the coil. The lower the inductance of the coil, and hence the lower its resistance, the greater will be the damping effect of the added resistance. The distinction is important; a fixed series resistance damps a small coil more than a large one, whereas a fixed parallel resistance has a greater effect on a large coil.

The resistance of a coil, or of a tuned circuit, depends very largely upon the frequency. With the ordinary small coil of some 160 μH, the equivalent series resistance may vary from some 25 ohms at 200 metres to perhaps 4 or 5 ohms at 550 metres. With decrease of frequency *r* drops ; but C, the capacitance necessary to tune the coil to the required frequency, rises, with the result that the dynamic resistance does not vary so greatly as the figures for *r* would suggest. In practice, the values for R vary over a range of about two to one over the medium-wave band. The high values for series resistance at low wavelengths are in the main due to dielectric losses, which, expressed as parallel resistance, are inversely proportional to frequency. A valve-holder that introduces $1\frac{1}{2}$ MΩ parallel resistance at 500 kc/s will introduce $\frac{1}{2}$ MΩ at 1,500 kc/s.

In conclusion, we see that the true representation of a tuned circuit as actually existing in a wireless set should include both series and parallel resistance, making a combination of Figs. 56 *a* and *b*. But, owing to the relationship existing between them, a circuit can be completely described at any one frequency by omitting either and making such an adjustment to the value of the other that it expresses the total loss of the circuit as a whole.

CHAPTER 7

THE TRIODE VALVE

62. Free Electrons

IN discussing the nature of an electric charge (Sec. 8) we saw that, if negative, it was due to an excess, or if positive, to a deficiency of electrons. We further saw that an electric current, such as might be obtained by connecting together two charged objects, consisted of a flow of these same electrons. In neither case, however, did the electron appear on its own, for it was always associated with matter.

In the thermionic valve we meet for the first time with electrons enjoying an entirely independent existence. Their source is the *cathode* of the valve, which is an electrically heated surface so prepared that when raised to a suitable temperature it emits into the vacuous space surrounding it a continual supply of electrons. These are too small and too light to feel appreciably the effects of gravity, and therefore do not tend to move in any particular direction unless urged by an electric field. In the absence of such a field they hover round the cathode, enclosing it in an electronic cloud known as the *space charge*.

Cathodes are of two types—*directly heated* and *indirectly heated*. The first, more usually known as a *filament*, consists of a fine wire heated by the passage through it of a current, the electron-emitting surface consisting of a film coated directly upon the filament wire itself. This type of valve is primarily used in battery-driven sets, though an occasional one finds its way into a mains-operated set in special cases. The indirectly heated cathode is a tube, usually of nickel, coated with the emitting material and heated by an independent filament, called the *heater*, enclosed within it. Since the cathode is insulated from the heater, three connections are necessary in this latter case as against the two that suffice when the filament serves also as the source of electrons, unless two are joined together or "commoned" as in certain rectifier valves.

In all essentials the two types of cathode work in the same way ; in dealing with valves we therefore propose to take the liberty of omitting the heater or filament circuit altogether after the first few diagrams, indicating the cathode by a single connection. The operation of a valve depends upon the emission from the cathode ; the means by which the cathode is heated to obtain this emission has little significance except in connection with the design of a complete receiver.

63. The Diode Valve

The simplest type of valve contains one electrode in addition to the cathode and is called a *diode*. This second electrode, the *anode*, will attract electrons to itself from the space-charge if it is made more positive than the cathode, so that a current can flow, through the valve, round a circuit such as that of Fig. 59. But if the battery is reversed, so that the anode is more negative than the cathode, the electrons are repelled towards their source, and, if made sufficiently negative, no current flows. The valve will therefore permit current to flow through it *in one direction only*, and it is from this property that its name is derived.

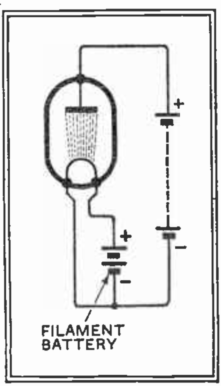


Fig. 59 : A directly heated (battery) diode valve

If the anode of a diode is slowly made more and more positive with respect to the cathode, as, for example, by moving upwards the slider of the potentiometer in Fig. 60, the attraction of the anode for the electrons is slowly augmented and the current increases. To each value of anode voltage E_a there corresponds some value of anode current I_a, and if the experiment is made and each pair of readings is recorded in the form of a dot on squared paper a curve like that of Fig. 61 is outlined.

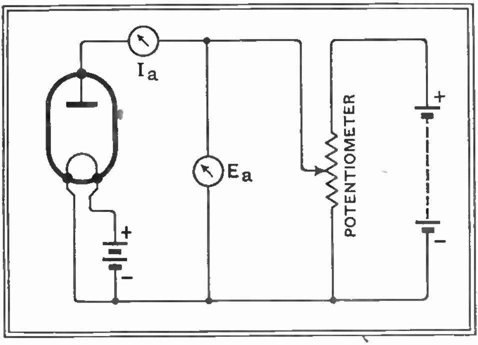


Fig. 60: Circuit for finding relation between anode voltage E_a and anode current I_a of a diode

The shape of the curve shows that the anode collects few electrons at low voltages, being unable to overcome the repelling effect of the space - charge. At the point A this is largely overcome and the increase in electron-flow with rising voltage becomes rapid and even. By the time the point C is reached the voltage is so high that electrons are reaching the anode practically as fast as the cathode can emit them ; a further rise in voltage only collects a few more strays, the current remaining almost constant from C to D.

At B an anode voltage of 100 volts drives through the valve a current of 4 mA ; it could therefore be replaced by a resistance of $100/0{\cdot}004 = 25{,}000$ ohms without altering the current flowing at this voltage. This value is therefore the equivalent D.C. resistance of the valve at this point. Examination of the curve

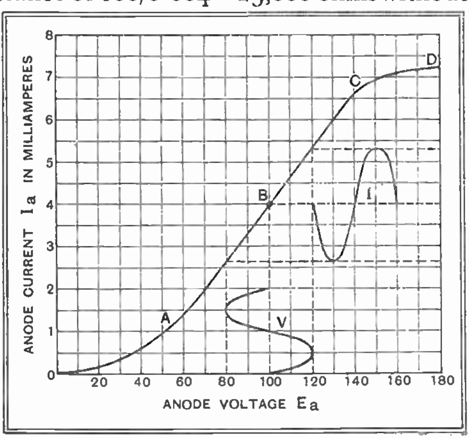


Fig. 61: Characteristic curve of a diode valve. The sine wave curves indicate the effect of superposing an alternating potential of 20 volts peak on the steady potential of 100 v. to which point B corresponds

will show that the equivalent D.C. resistance of the valve *depends upon the voltage applied* ; to drive 1 mA, for example, needs 53 volts, which leads to R = 53/0·001 =53,000 ohms.

64. Anode A.C. Resistance (r_a)

One may, however, deduce the resistance of the valve in another way. Over the straight-line portion of the curve, round about B, an increase of 30 anode volts brings about an increase in anode current of 2 mA. The resistance over this region of the curve would therefore appear to be 30/0·002 = 15,000 ohms. This resistance is effective towards current-variations within the range A to C ; if, for example, a steady anode voltage of 100 volts were applied (point B) and then an alternating voltage of peak value 20 volts were superposed on this, the resulting alternating current through the valve, as the curves on Fig. 61 show, would be 1·33 mA peak. Based on this, the resistance, as before, comes out to 20/1·33 = 15,000 ohms. * Thus the resistance derived in this way is that offered to an alternating voltage superposed on the steady anode voltage ; it is therefore called the *Anode A.C. resistance* of the valve. Its importance in wireless technique is so great that it has had the special symbol r_a allotted to it by common consent of wireless engineers. It is also, but not so correctly, called the " impedance " of the valve. The number of steps up in current for one step up in voltage is the *slope* of the curve. But r_a is the number of voltage steps for one current step, so it is 1 divided by the slope, or, as it is called, the reciprocal of the slope, at the particular point selected on it. As can be seen from Fig. 61, a steep slope means a low anode A.C. resistance.

The equivalent D.C. resistance of a valve is a quantity seldom used or mentioned ; it was discussed here only for the sake of bringing into prominence the strictly A.C. meaning of the valve's impedance.

65. The Triode Valve

The diode valve has a very restricted field of use in that it can be used for rectification only ; it will not provide

* By now the reader should have noticed that volts, *milli*amps, and *thousands of* ohms form a self-consistent system to which Ohm's Law applies. This offers a short cut in many wireless calculations.

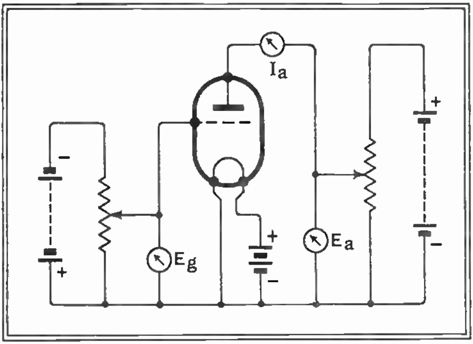


Fig. 62 : Circuit for taking characteristic curves, as in Fig. 63 or 64, of triode valves

amplification. If a mesh of fine wire is inserted in the valve between cathode and anode in such a way that before they can get to the anode all the electrons emitted from the cathode have to pass through the meshes of this extra electrode a much fuller control of the electron-current becomes possible.

It is fairly evident that if this new electrode, the *grid*, is made positive it will tend to speed up the electrons on their way through its meshes to the anode; if, on the other hand, it is made negative it will tend to repel them back towards the cathode. In Fig. 63 are shown four

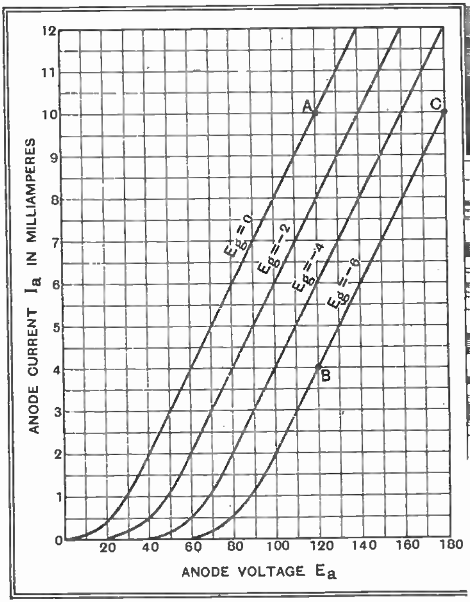


Fig. 63 : Characteristic curves of triode valve, each showing change of anode current with change of anode voltage

curves of a three-electrode valve, or *triode*, for comparison with the exactly analogous curve of the diode (Fig. 61). Each of these curves was taken with a fixed grid voltage which is indicated against each curve. It is to be noticed that this voltage, like all others connected with a valve, is reckoned *from the cathode as zero.* If, therefore, the cathode of a valve is made two volts positive with respect to earth, while the grid is connected back to earth, it is correct to describe the grid as "two volts negative," the words "with respect to the cathode" being understood. In a directly heated valve voltages are reckoned from the *negative end* of the filament.

66. Amplification Factor (μ)

Except for a successive displacement to the right as the grid is made more negative, these curves are practically identical. This means that while a negative grid voltage reduces the anode current in the way described, this reduction can be counterbalanced by a suitable increase in anode voltage. In the case of the valve of which curves are shown, an anode current of 10 mA can be produced by an anode voltage of 120 if the grid is held at zero potential ($E_g = 0$). This is indicated by the point A. If the grid is now made 6 volts negative the current drops to 4 mA (point B), but can be brought up again to its original value by increasing the anode voltage to 180 v. (point C).

A change of 6 volts at the grid can thus be compensated for by a change of 60 volts, or ten times as much, at the anode. For reasons that will presently appear, this ratio of 10 to 1 is called the *amplification factor* of the valve, and is denoted by the Greek letter "μ" (mu). The letter *m* is also sometimes used.

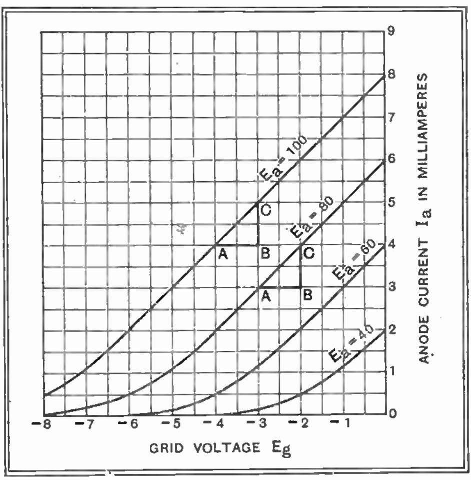
As in the case of the diode, the anode resistance of the valve, by which is again meant the resistance it offers to the passage through it of a small alternating current when a small alternating voltage is superposed on some steady anode voltage, can be read off from the curves. All four curves of Fig. 63 will give the same value over their upper portions, since they all have the same inclination ; over the lower parts, where the steepness varies from point to point, a whole range of values for the anode resistance exists. Over the straight-line portions of the curves this resistance is 10,000 ohms, as can be seen from the fact that the anode voltage must change by 10 to alter the anode current by 1 mA.

67. Mutual Conductance or Slope (g_m)

We have already seen that 1 volt on the grid is equivalent to 10 volts on the anode ; a change of 1 volt at the grid will, therefore, also cause a change in the anode current of 1 mA. This can also be read directly from the curves by observing that at $E_a = 100$, the anode current for $E_g = 0$ and $E_g = -2$ are 8 and 6 mA respectively, again a change of 1 mA for each one-volt change on the grid.

The response of the anode current of a valve to changes in voltage at the grid is the main index of the control that the grid exercises over the electron-stream through the valve. It is expressed in terms of *milliamperes* (of anode-current change) *per volt* (of change at the grid), and is called the *mutual conductance* (symbol g_m). It is related to μ and r_a by the simple equation $g_m/1000 = \mu/r_a$, or, if g_m is in *amperes* per volt, $g_m = \mu/r_a$; the derivation of which should be evident if the meanings of the symbols are considered. The magnitude of g_m is more clearly

Fig. 64 : Changes of anode current corresponding to variations of grid voltage



shown by valve-curves in which anode current is plotted, for a fixed anode voltage, against grid voltage. Some data from Fig. 63 are replotted in this form in Fig. 64, where the lines BC represent the anode-current change brought about by a change AB in grid voltage. The ratio BC/AB is very evidently the mutual conductance of the valve in milliamperes per volt. Since this ratio also defines the slope of the curve, it has become quite common to refer to g_m as the "*slope*" of the valve. But it must be clearly understood that it is the slope of this particular valve characteristic (anode current against grid voltage) that is meant. The slope of the anode-current—anode-volts curve (Fig. 61) is obviously different, and, in fact, is the anode conductance $1/r_a$.

68. Alternating Voltage on the Grid

The "characteristic curves" of a valve, whether the anode current is shown plotted against anode voltage or against grid voltage, do not give complete information as to how the valve will behave in the set. They do, however, provide the necessary data from which its performance can be determined.

In Fig. 65 we have a set of $E_g - I_a$ curves for a typical triode of the medium-impedance class. As the slope of the curves shows, its mutual conductance g_m is about $3\frac{1}{2}$ to 4 mA per volt for anode currents in excess of about 4 mA, but less for lower currents. Suppose that, as suggested in the inset to that figure, we apply a small alternating voltage V_g to the grid of the valve, what will the anode current do? If the batteries supplying anode and grid give 200 and $-2\frac{1}{2}$ volts respectively, the anode current will set itself at about $5\frac{3}{4}$ mA—point A on the uppermost curve.

If the alternating voltage applied to the grid has a peak value of 0·5 volt, the total voltage on the grid will swing between -3 and -2 volts, alternate half-cycles adding to or subtracting from the initial (negative) grid voltage. The anode current will swing in sympathy with the changes in grid voltage, the points B and C marking the limits of the swing of both. The current, swinging between $7\frac{1}{2}$ and 4 mA, is reduced by $1\frac{3}{4}$ mA on the negative half-cycle and increased by the same amount on the positive one. The whole is therefore equivalent to the original steady current with an alternating current of $1\frac{3}{4}$ mA peak superposed on it.

The development of an alternating anode current in response to the signal is not, however, enough. The next valve in the chain will require an alternating *voltage* to operate it. To develop this voltage we have to put an impedance of some kind in the anode circuit of our valve, so that the alternating current through it shall develop the alternating voltage we want.

69. The Load Impedance

In principle this is simple (see inset to Fig. 66), but it brings a complication in its train. The circuit referred to shows very clearly that the alternating voltage is developed

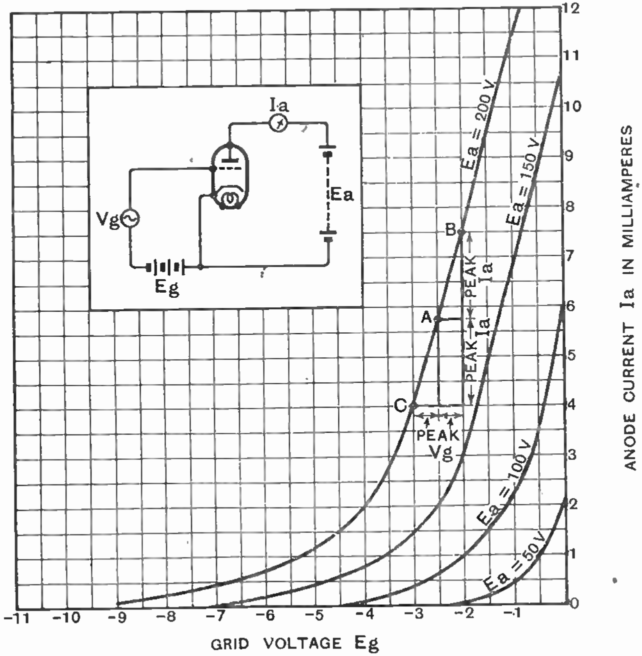


Fig. 65: Anode-current grid-voltage curves of a medium-impedance indirectly heated triode. The alternating anode current caused by an alternating grid-voltage V_g can be read from the curves

actually on the anode of the valve ; we can therefore no longer assume, as in discussing Fig. 65, that the anode voltage is constant. Instead, it rises and falls with the alternations of the signal.

To find out what happens when grid and anode voltages vary together in this way we again have recourse to the valve curves, using this time the E_a—I_a curves of Fig. 66, each of which refers to a definite fixed grid-voltage as indicated against the curves themselves.

The inset to Fig. 66 indicates that the battery supplies 240 v. to the anode circuit as a whole. The steady anode current through the anode load resistance R will drop across this resistance some portion of the applied voltage ; at the anode itself the voltage will therefore be less than 240 v. For any given value of R we can plot voltage-at-anode

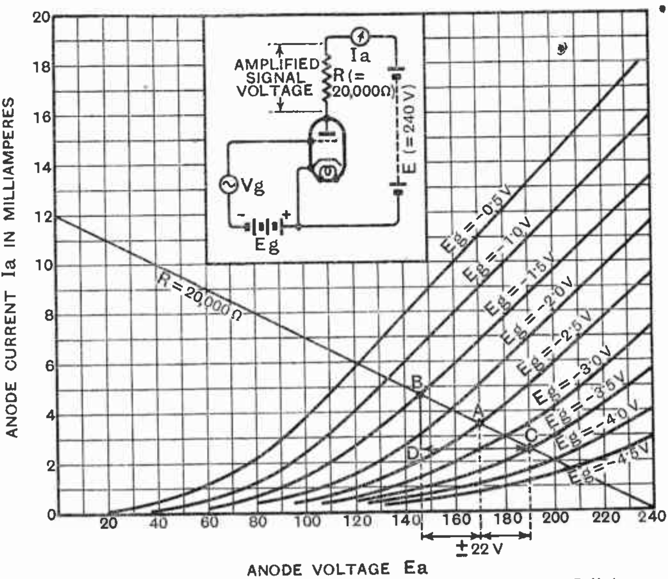


Fig 66 : E_a—I_a, curves of the valve of Fig. 65. The line "R = 20,000 Ω" drawn across the curves gives, in conjunction with the curves themselves, full data as to the performance of the valve with this value of load resistance and an anode battery of 240 v.

against anode-current; if $R = 20{,}000\ \Omega$, there will be lost across it 20 volts for every milliamp. flowing, and the voltage at the anode will be reduced below 240 v. by this amount. Thus the anode voltage will be 200 if $I_a = 2$ mA, 160 if $I_a = 4$ mA, and so on. Plotting these points gives us the line " $R = 20{,}000\ \Omega$ " of Fig. 66.

The reason why it slopes in the opposite direction to the valve curves is that it represents the ratio of *loss* of anode voltage to increase in anode current.

From the way in which this line has been derived it is evident that every possible combination of I_a and E_a is expressed by some point along its length. Each of the valve curves across which it falls indicates the combinations of E_a and I_a that are possible for the particular value of grid-bias indicated against that curve. It follows that if we set the bias at $-2\frac{1}{2}$ v. with the load resistance in circuit and connected to 240 v. as shown, I_a and E_a will be indicated by the point A, since the *working point* has to fulfil the double conditions of lying on both straight line and curve. For any other value of bias the anode current and voltage would equally take the values shown by the intersection of the *load-line* with the corresponding curve.

If the grid-voltage of the valve is slowly increased from $-\frac{1}{2}$ v. towards $-4\frac{1}{2}$ v. the anode current will fall, as in Fig. 65, but the fall will be slower than in that figure since the anode voltage will rise as the current drops, as shown by the intersections of the load-line with the curves for successive values of bias.

Picking out and plotting the values of I_a for these intersections we get the heavy-line curve of Fig. 67. This is the *dynamic* or *working* characteristic of the valve when used in the particular circuit we are considering. Comparing it with the ordinary or *static* curves shown in dotted lines in Fig. 67, we see that it has a much lower slope, thus representing a lower mutual conductance, than these. The working value of the slope of the valve-resistance combination is approximately 1 mA/v., as compared with nearly three times this value for the valve alone.

70. Dynamic Characteristic

If we remember that $g_m = \mu/r_a$, the reason for the reduction in slope through the inclusion of the resistance becomes evident ; a change in E_g of 1 volt, equivalent to a change in E_a of μ volts, now has to produce its anode current change through r_a and R in series, instead of through r_a alone.

Hence the working slope of the valve-resistance combination is $\mu/(R + r_a)$.

At the working-point A in Fig. 66 the slope of the curve shows that $r_a = 14{,}400\,\Omega$, while the horizontal spacing between curves shows that $\mu = 39$. This gives, for the valve alone, $g_m = 39/14{,}400 = 2{\cdot}71$ mA/volt. By reading directly from the dotted curves of Fig. 67 at an anode current of 3·45 mA (the current at A) we get the figure 2·65 mA/volt, which is as close an agreement as free-hand curves are likely to give.

For the working slope, calculation gives $\mu/(R + r_a) = 39/(14{,}400 + 20{,}000) = 39/34{,}400 = 1{\cdot}13$ mA/volt. Direct check from the heavy curve of Fig. 67 gives 1·10 mA/volt. The theory upon which the calculation was based is thus confirmed.

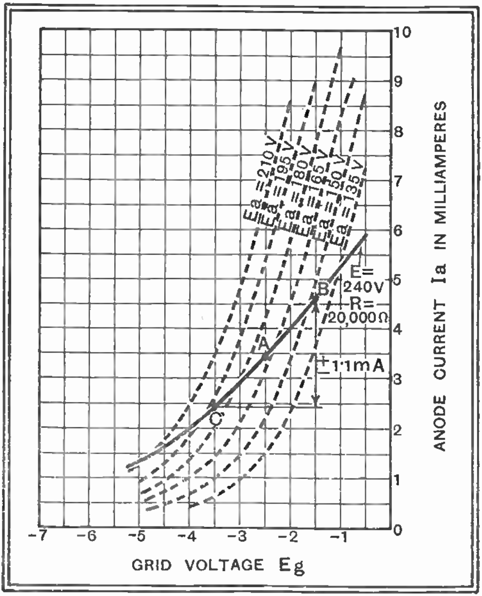
71. The Triode as Amplifier

This study of the effect of a resistance in the anode circuit upon the characteristic curves of a valve has brought us to a point from which the behaviour of the valve as an amplifier is immediately apparent. First, the calculation. Let us start up the A.C. generator V_g of Fig. 66 and assume it delivers 1 volt to the grid of the valve. This is equivalent, as we know, to introducing μ volts, or in this particular case 39 volts, of A.C. into the anode circuit. Applied to r_a and R in series this will produce a current of $39/34{,}400 = 1{\cdot}13$ mA of A.C. This, flowing through R ($= 20{,}000\,\Omega$) will cause a potential-difference of $1{\cdot}13 \times 20 = 22{\cdot}6$ volts.

If one volt of signal applied to the grid produces 22·6 volts across R, the amplification of the whole *stage* (valve *plus* resistance) is 22·6 times.

From the way in which we have arrived at this figure it is not very difficult to obtain the formula for calculating the amplification, usually denoted by A, given by any stage : $A = \mu R/(R + r_a)$. Now we will check this directly from

Fig. 67: Dynamic characteristic of the valve of Figs. 65 and 66 used in the circuit inset on the latter Figure. The dotted lines are ordinary Eg-Ia curves, as in Fig. 65. Note the much decreased slope of the dynamic characteristic and that a different dynamic characteristic could be plotted for every combination of load resistance and battery voltage



the curves. In Fig. 66, the point A lies on the curve $E_g = -2\frac{1}{2}$ v. If we superpose on this a signal of 1 volt peak the grid will swing between $-3\frac{1}{2}$ and $-1\frac{1}{2}$ v. Anode current and anode voltage will then swing over the straight line BC, which covers a voltage-swing of $\pm$ 22 v. Thus a one-volt swing on the grid results in a 22-volt swing on the anode, as calculated.

Either on this figure or on Fig. 67 we can see the alternating anode current. C lies at 2·45 mA, B at 4·65 mA, a swing of $\pm$ 1·1 mA., as calculated, round the initial steady value at A.

For most purposes, all the information likely to be of use can be read off at once from a set of curves such as those of Fig. 66 in conjunction with the necessary load-line.

72. The Effect of Load on Amplification

The formula for amplification that we have used, $A = \mu R/(R + r_a)$, shows at once that by making R so large that r_a is negligible in comparison with it, the amplification given by the stage will rise towards a theoretical maximum equal to μ, which supplies the reason for calling this quantity the "amplification factor" of the valve.

The same result can be had graphically by considering Fig. 66. For a higher value of R than 20,000 Ω, the line would be more nearly horizontal; assuming the working point A retained, it would cut the axis $I_a = 0$ at a higher voltage. Imagining the line pivoted round A till it becomes horizontal (R infinitely high) we arrive at a diagram in which the anode current remains constant while the anode voltage changes, and so leads to μ as the stage-gain.

Conversely, the effect of a lower load can be studied by tipping the load-line towards the vertical; evidently a lower stage-gain would be produced until, in the limit, the line becomes vertical (R = 0) and there is no change in voltage at the anode in response to the signal.

Fig. 68 shows the values of stage-gain obtained with various anode loads for a valve in which $r_a = 14{,}400$ Ω and $\mu = 39$ *. It should be added that such a curve does

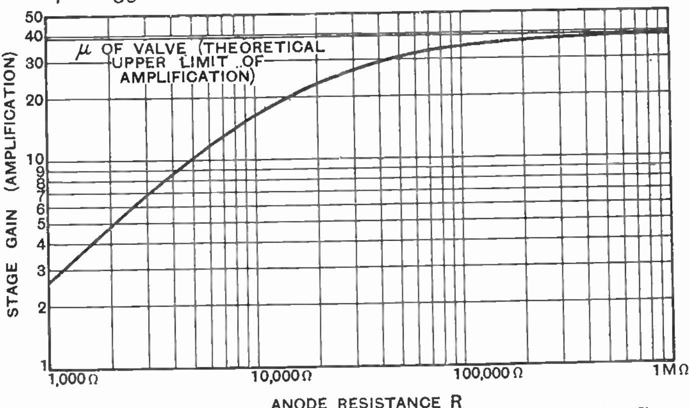


Fig. 68: Amplification of stage with different values of load resistance R. The valve, throughout, is supposed to have $\mu = 39$, $r_a \doteq 14.400$ ohms

* *Logarithmic* scales are used for R and A in order to cover a wide range without compressing the most interesting part of the curve on the extreme left—a useful dodge that will be employed in other parts of the book. It is a scale in which equal distances represent equal *multiplications* instead of equal *additions*. For example, one-third of the whole R scale represents a ten-fold increase no matter where that third is taken.

not take into account the fact that, unless very high anode voltages (external to R) are used, r_a will not stay constant as assumed, but will rise with the falling anode current. True values can only be obtained by analysis of actual curves on the lines of Fig. 66.

73. Power in Grid and Anode Circuits

The observant and enquiring reader may have wondered why the grid of the valve has been shown as always negative, and never positive, with respect to the cathode. The reason is bound up with the desire to expend as little power as possible in the grid circuit. If the grid were allowed to run positive it would collect electrons instead of making them all pass through its meshes, and a current would then flow round the grid circuit, absorbing power from the generator. Since this may be, in practice, a tuned circuit of high dynamic resistance, this absorption of power would have markedly ill effects in reducing the voltage across it and in decreasing the effective selectivity.

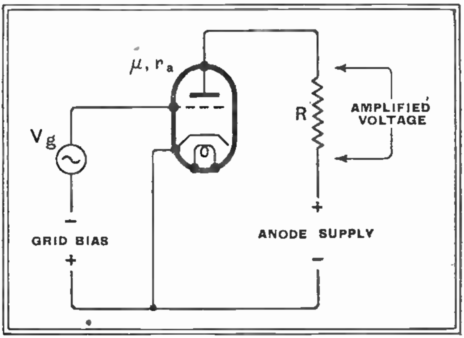
Provided no current flows in the grid circuit we may, for the moment, regard the valve as absorbing no power in that circuit. This condition will always be fulfilled if the initial negative voltage, known as *grid bias*, applied by the battery, makes the grid negative enough to prevent the flow of grid current even at the peak of the positive half-cycle of signal voltage. In general, the bias required is equal to, or a volt or so greater than, the peak of the signal that the valve has to accept.

In spite of the fact that the power consumed in the grid circuit is negligibly small, alternating voltages applied to the grid can release an appreciable amount of A.C. power in the anode circuit. We have just discussed a case in which 1·1 mA peak of A.C. developed 22 volts across a resistance, making $\frac{1}{2}(22 \times 1{\cdot}1) = 12{\cdot}1$ milliwatts of power. This power is, of course, derived from the anode battery, which is continuously supplying 3·45 mA at a total of 240 volts, which is 829 milliwatts.

Behind all the curves and calculations there lies the simple basic fact that the valve is able to convert the D.C. power from the battery into A.C. power in response to a practically wattless A.C. driving-voltage on its grid. It is to this conversion, at bottom, that it owes its ability to amplify.

74. Six Important Points

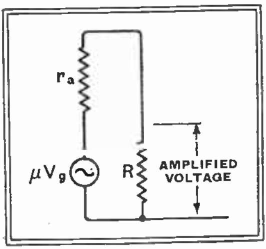
For reference, and as a summary of this chapter, we will tabulate the most important points about the triode valve.



(a)

Fig. 69: (a) A stage of amplification, and (b) a diagrammatic representation of the anode circuit only, the signal voltage V_g on the grid being replaced by its equivalent μV_g volts, in series with the valve's own anode-cathode resistance r_a. Only Ohm's Law is needed to show that the voltage on R is $\frac{\mu R}{R + r_a}$ times V_g

(1) One volt at the grid controls the anode current to the same extent as μ volts at the anode; μ is the amplification factor of the valve.



(b)

(2) Towards A.C. the valve has an anode resistance r_a depending for its exact value upon the steady voltages applied.

(3) The control of anode current by grid voltage is given by the ratio μ/r_a, known as the mutual conductance or slope, g_m.

(4) As a corollary to the above, it follows that a valve can be represented as a resistance r_a in series with a generator the voltage of which is μ times the alternating voltage applied to the grid. This representation (Fig. 69) takes no account whatever of steady voltages and currents, except through their influence in determining μ and r_a.

(5) The amplification given by a valve in conjunction with its load resistance R is $A = \frac{\mu R}{R + r_a}$, as Fig. 69 clearly shows.

(6) Since r_a, μ, and g_m all depend, to a greater or lesser extent, on actual operating voltages, all *detailed* study of a valve's behaviour must be made by drawing load-lines across its actual curves, as in Fig. 66. Deductions such as Fig. 67 can then be made from these.

(7) The more elaborate valves which we shall discuss later are really only improved versions of the triode. In consequence, these six points cover about 90 per cent. of the philosophy of these more complicated structures.

CHAPTER 8

THE NATURE OF THE RECEIVED SIGNAL

75. The Raw Material of Reception

IN Chapter 1 there was a brief but necessarily incomplete description of the signal picked up by a receiving aerial; it was described as "a wireless wave that bears upon it, in the form of variations in strength, the impress of the currents derived from the microphone in the studio".

The whole complex assembly of valves and circuits making up a wireless set is designed with the one aim of amplifying this signal and converting it into new and more useful forms; we cannot even begin to discuss the set itself until we have expanded this too-concise description into something much more explicit and detailed.

76. The Simple Carrier Wave

If we imagine that radio-frequency currents are generated, in some unspecified way, in the tuned circuit L_1C_1 of Fig. 70, then if L_1 is coupled to L_2, as the diagram suggests, similar currents will appear in the latter coil. If this is connected between earth and an aerial, the capacitance between the aerial and the ground beneath it can be used to tune L_2 to the frequency of the current, as indicated by the dotted condenser

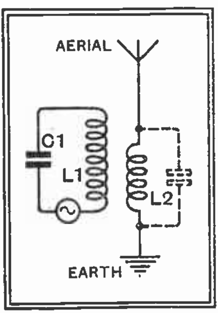


Fig. 70: Schematic diagram showing how a wireless wave originates at the transmitting station. The coil L_2 is tuned by the capacitance (shown in dotted lines) existing between the aerial wire and the ground beneath it

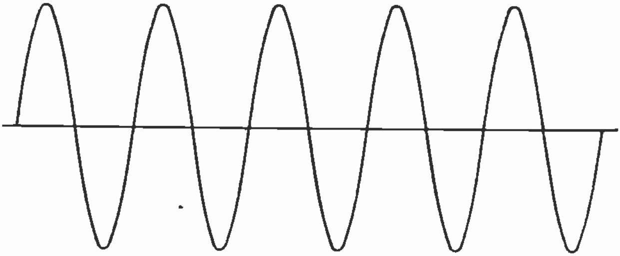


Fig. 71 : Curve representing rise and fall of current in L_1, C_1 (Fig. 70), rise and fall of voltage on aerial, or rise and fall in strength of field due to radiation. If f = 1,000 kc/s (λ = 300 m.) the unit of time is one-millionth of a second

which represents the aerial-earth capacitance. This means that the aerial will be charged and discharged at radio frequency as was the condenser in the simpler circuit described in Sec. 30.

We have already had occasion (Sec. 31) to represent the radio-frequency current in an oscillating circuit by a sine-wave, as in Fig. 71. If this current is conveyed to the aerial in some such way as suggested in Fig. 70, the aerial and earth system will send out into space an electromagnetic wave consisting of varying electric and magnetic fields which travel outwards from the aerial with the speed of light. (Chap. 19). At any point to which these fields reach on their travels, their intensity varies with time in exactly the same manner as does the current in the aerial. Fig. 71 will therefore serve to represent this wave, although the curve is in no sense a physical picture of it ; it is simply a record of the way in which the intensity of the field varies with time.

If a continuous wave of this type were sent out from a transmitter it could convey no more information than could a steady beam of light from a lighthouse. Lighthouses are accustomed to announce their identity to the navigator by periodic rhythmic interruption of their light, sending out in this way a sort of "call-sign" of long and short flashes. In just the same way a wireless transmitter can convey messages by periodically interrupting its wave, breaking it up into the short and long bursts of transmission that represent the dots and dashes of the Morse code.

To convey speech and music, something more elaborate than simple interruption of the transmission is required, though the continuous *carrier wave* is still the basis of this more advanced type of transmission.

77. Modulating the Carrier

We will imagine that it is desired to transmit a pure note of 1,000 c/s, available in the form of an electric current derived, in the first place, from the microphone before which the note is being sounded. This current will also have a form like that of Fig. 71, but the time-scale will be profoundly different from that used when the curve represents a radio wave. If the wave corresponds to 300 metres, or a frequency of 1,000,000 c/s, each audio cycle will extend over a thousand radio cycles. To enable the musical note to be conveyed by the carrier wave these two oscillations have to be combined to make a single whole.

In Fig. 72 *a* is depicted the "wave-form" of a radio-frequency carrier wave, while *b* shows, to the same scale, the musical note which we wish to combine with it. At first sight it might seem that it would be sufficient to add the two currents together and allow them both to flow in the aerial. Such a mode of combining them results in the wave-form shown in full-line in Fig. 72 *c*. Examination of this figure will show that the two currents, although they are flowing simultaneously in the same circuit, are still independent, the whole consisting of the original radio-frequency current oscillating round a zero voltage which moves slowly up and down at the frequency of *b*. The dotted curve shows the new zero voltage. Successive peaks of the radio-frequency voltage are still exactly alike, as they were in *a*.

In view of the known fact that an aerial will not radiate an audio-frequency voltage to any appreciable extent, it is clear that if an attempt were made to send out *c* as a signal, the radio-frequency component would set up its usual wave, as at *a*, in which the audio-frequency component would not be represented.

It is clear, therefore, that simple *addition* of the currents will not provide us with a resultant current of a suitable type for radiation ; we will therefore try *multiplication*.

Let us suppose tha· the amplitude of the radio-frequency current in the aerial depends upon the D.C. voltage used to drive some part of the apparatus generating the oscillation. The height of curve *a* might then be 1 amp. if a

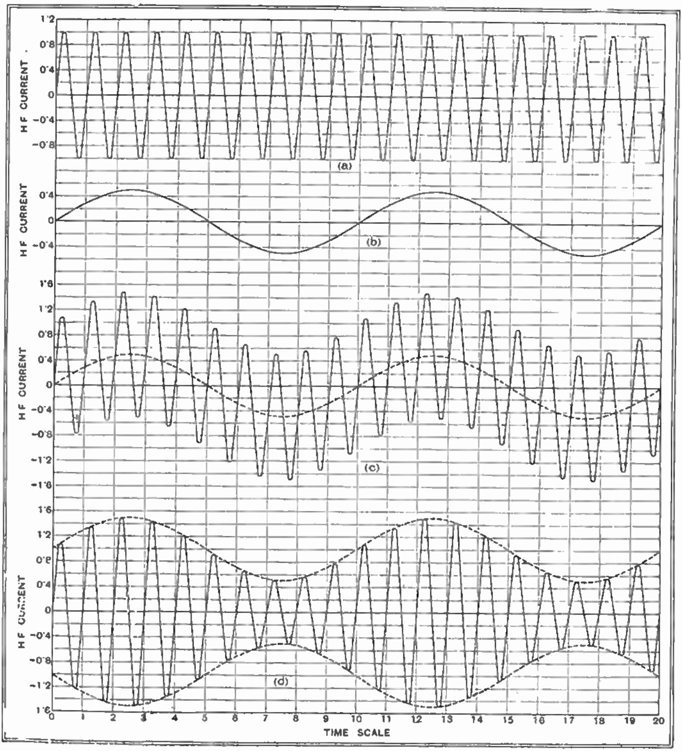


Fig. 72: Diagrams *a* and *b* show a radio-frequency current and a musical note to be combined with it for transmission. Mere addition of the currents results in *c*, in which the currents remain separate so that only the R.F. component would be radiated. Diagram *d* shows the modulated carrier resulting from multiplying the curves as described in the text : it is radiated complete from the aerial

500-volt battery were used, but might rise to 1·5 or drop to 0·5 amps. if the battery were suitably increased or decreased in voltage. We might now introduce the audio-frequency voltage we desire to transmit *in series* with this imaginary battery; then the total voltage reaching the R.F. generator would swing about its mean value, the audio-voltage alternately adding to and subtracting from the battery voltage. In consequence the amplitude of the radio-frequency output from the generator would also rise and fall, this rise and fall being strictly in time with the audio-frequency voltage we wish to transmit.

The result of this more elaborate means of combining the two curves, which amounts to multiplication of the one by the other, is shown at *d* in Fig. 72, where it will be seen that the amplitude of the radio-frequency current is now actually changing at audio-frequency. Except as an impress on the total amplitude of swing the audio-frequency current has disappeared; it is now represented by the *envelope* (dotted) of the curve as a whole.

A curve such as *d* represents a *modulated* high-frequency current or voltage. It is fairly evident that if this is allowed to flow in an aerial the radiated wave will follow, in its rise and fall, the rise and fall of the current, since the whole is now a radio-frequency phenomenon.

The observant reader will have noticed one important inaccuracy in the diagram; it does not bring out clearly enough the enormous difference in frequency between the carrier and the modulation. If, as suggested, *b* shows a 1,000-cycle (1 kc/s) note, *a* represents a 10-kc/s carrier, having a wavelength of 30,000 metres. To show a 1,000-kc/s (300-metre) carrier in its correct relationship to *b* there should be 100 complete radio-frequency cycles in the place of every one shown. A little imagination must therefore be applied to Fig. 72 before it can give a correct impression of a normal broadcast wave.

Even so, *d* represents nothing more exciting than a tuning-note; for music or speech the form of *b* is extremely complex, and this complexity is faithfully represented in the envelope of the modulated carrier *d*. Nevertheless, the diagram gives a very fair mental picture of the modulated carrier which flows, as a current, in the transmitting aerial, and is radiated outwards through space as a wireless wave.

In Fig. 72 and elsewhere "HF" (high-frequency) is used to refer to the frequency of the carrier wave, following a general custom. But whereas 5,000 c/s is "high frequency" compared with, say, 50 c/s, it is not suitable for a carrier wave. It is less confusing, therefore, to use the term "radio frequency" (RF) in place of HF to mean a frequency that is suitable for radiating; and "audio frequency" (AF) in place of LF to mean a frequency that is audible. As there are a number of these confusing equivalent terms in use, a list of them is given at the end of the book.

78. Depth of Modulation

Since our receiver will be so designed that the carrier wave itself, in the intervals of modulation, (between items in the programme) gives rise to no sound in the loudspeaker, it is evident that a curve such as that of Fig. 72 *d* represents a note of some definite loudness, the loudness depending on the amount by which the radio-frequency peaks rise and fall above and below their mean value. The amount of this rise and fall in proportion to the normal amplitude of the carrier wave is spoken of as the *depth of modulation.*

For distortionless transmission the increase and decrease in carrier amplitude that correspond to positive and nega-

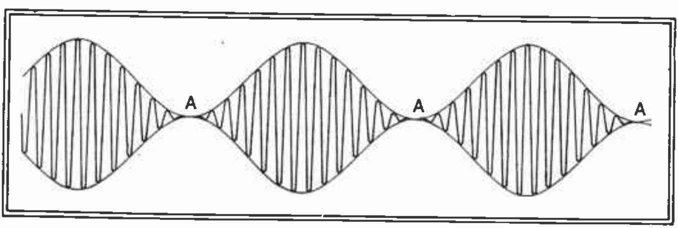


Fig. 73: Carrier-wave modulated to a depth of 100%. At its minima (points A) the R.F. current just drops to zero; any attempt at still deeper modulation results in a series of separate bursts of current, and the envelope no longer has the form of the modulating wave

tive half-cycles of the modulating voltage must be equal. It is evident that the maximum possible decrease in carrier-amplitude is found when the modulation reduces the carrier so far that it just, and only just, ceases at the exact moment of minimum amplitude, as shown in Fig. 73. At its maximum it will then rise to double its steady value. Any attempt to make the maximum higher than this will result in the carrier actually ceasing for an appreciable period at each minimum ; over this interval the envelope of the carrier-amplitude can no longer represent the envelope of the modulating voltage, and there will be distortion.

When the carrier has its maximum swing, from zero to double its mean value, it is said to be modulated *to a depth of* 100 *per cent.* In general, the maximum rise in amplitude, expressed as a percentage of the mean, is taken as the measure of modulation depth. Thus a rise from 1 volt to 1·5 volt corresponds to 50 per cent. modulation, a rise to 1·4 volt to 40 per cent., and so for other values.

In transmitting a musical programme, variations in loudness of the received music are produced by variations in modulation-depth, these producing corresponding changes in the audio-frequency output from the receiver.

CHAPTER 9

DETECTION

79. The Need for Detection

AT the receiving aerial, the modulated carrier-wave sets up currents which, apart from any distortion they may have suffered in their journey through space, are an exact duplicate in miniature of the currents in the aerial of the transmitter. By some simple circuit, such as that of Fig. 74 (compare Fig. 70), they can be collected and caused to flow, in magnified form, round a tuned circuit. The function of even the simplest receiver is to convert these electric currents into sound so that the programme may be enjoyed.

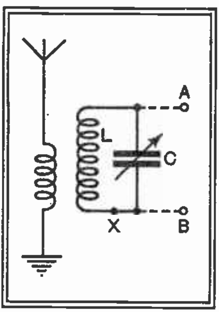


Fig. 74 : Showing how the modulated carrier is "collected" by the aerial in the form of currents of the same wave-form and passed to the tuned circuit LC as the first stage in reception

If earphones were connected to the circuit, either by inserting them at X to allow the circulating current to flow through them or by joining them across A and B so that the voltage on C would drive a current through them, no sound would be heard. The reason for this can readily be appreciated by considering Fig. 75 *a*, which repeats the diagram of the modulated carrier. Any two consecutive half-cycles of the current are approximately equal (in a practical case, much more nearly equal than in the diagram) and so neutralize each other so far as the earphone diaphragm is concerned, it being understood that this cannot possibly vibrate at anything like such a high frequency as that of the carrier. The average current through the phones, even

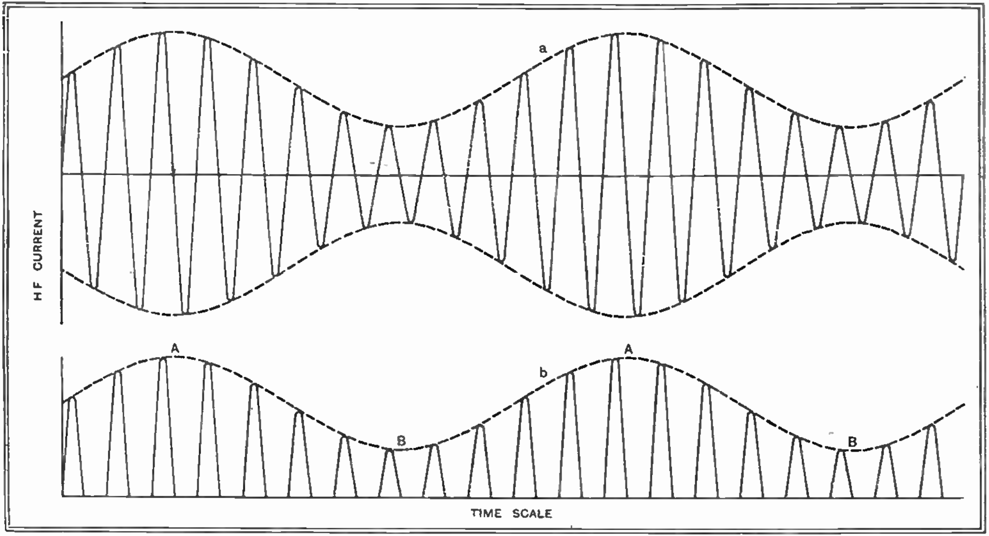


Fig. 75 : Modulated R.F. current (diagram *a*). Over any period of time appreciably greater than an R.F. cycle, the average current is zero, and so is inaudible in earphones. The same current, rectified, is shown in diagram *b*. The average current now rises and falls at modulation-frequency and can now be heard in earphones

measured over an interval as short as a ten-thousandth of a second, is therefore zero.

But if we could find some means of wiping out one-half of the wave, so that it took on the form shown at *b*, we should have a current to which the phones could respond, for the average current would then be greater at A than at B. While unable to follow the carrier-frequency alternations individually, the phone diaphragm would then rise and fall at the rate of their variation in amplitude. Since these variations are due to the audio-frequency note modulating the carrier, it is this note, which we want, that would be heard.

The process of suppressing half of a complete wave, thus converting alternations of current into a series of pulses of unidirectional current, is called by the general term *rectification*. The particular case of rectifying a modulated carrier in such a way as to reveal the modulation is known in this country as *detection*, and in America as *demodulation*. It can be performed by any device which conducts current, or responds to a voltage, in one direction only, or, less perfectly, by any device which has a lower resistance to currents, or a greater response to voltages, in one direction than in the other.

80. Types of Detectors

A perfect detector would be one that had no resistance to current flowing in one direction, and an infinite resistance to current in the opposite direction. It would, in fact, be equivalent to a switch, completing the circuit during all the positive half-cycles in Fig. 75*a*, thus enabling them to be passed completely, as in *b*; and interrupting it during all the negative half-cycles, suppressing them completely, as also shown in *b*. For very low frequency alternating currents, such as 50 c/s, it is possible to construct such a switch (known as a vibrator rectifier); but any device involving mechanical motion is hardly practicable at radio frequencies.

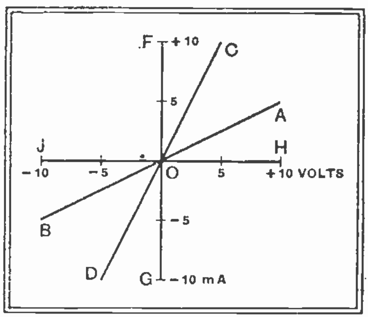
There are a number of minerals that have a lower resistance to current in one direction than the other, and they were at one time very largely employed as detectors in *crystal* receivers.

More recently the manufactured copper oxide rectifiers used extensively for low frequency rectification have been adapted for radio frequencies and are sold under the trade name "Westector." But the effect most widely used is the one already met in Sec. 63—the one-way movement of electrons in a vacuum—of which the thermionic valve is the embodiment.

81. Detector Characteristics

To understand the various available types of detectors it is necessary to make use of characteristic curves, in which the relation is shown between applied voltage and the resulting current. These have already been used in

Fig. 76: Characteristic curves of various "linear" resistances. The line COD, by Ohm's Law, is seen to represent 500 ohms. If the resistance to negative voltages is different from that to positive, as shown for example by the line COB, the result is a partial rectifier



connection with valves (e.g., Fig. 63), but it may be as well to make sure that their significance is grasped. In Fig. 76, the current flowing in a conductor is plotted against applied voltage. In an ordinary conductor having a definite unaltering resistance, the graph is a straight line; for example, AOB, which, by Ohm's Law, can be seen to represent 2,000 ohms. The line COD represents 500 ohms.

The steeper the slope, the lower the resistance, as we saw in Sec. 64. The line FOG represents zero resistance, and HOJ infinite resistance; so FOJ would be the graph of a perfect rectifier. COB would be an example of a partial rectifier—a resistance of 500 ohms to positive currents and 2000 ohms to negative currents. Now apply an alternating voltage to a circuit consisting of any of the foregoing imaginary resistances in series with a fixed resistance of 1000 ohms (Fig. 77*a*). The voltage across the 1,000 ohms—call it the output voltage—is equal to the applied voltage only if R is zero. Both voltages are then indicated by the full line in Fig. 77*b*. If R were 500 ohms (COD), the output would be represented by the dotted line CD, and if 2,000 ohms by the dotted line AB.

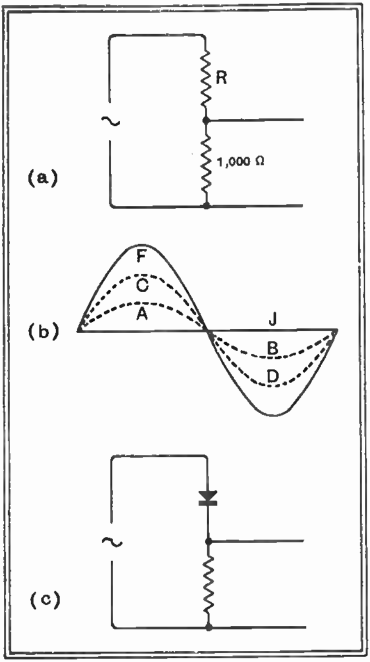


Fig. 77 : The results of applying a sine wave voltage to resistances represented by Fig. 76, in the circuit shown here at a, are indicated by b, which is lettered to correspond with the various characteristics in Fig. 76. When R is a rectifier, it is represented by the symbol shown in c

82. Load Resistance and Output

If now a rectifier is substituted for R, the symbol being as shown in Fig. 77*c*, the positive and negative output voltages are unequal. For example, the perfect rectifier (FOJ) gives the output FJ in *b* ; and the partial rectifier gives CB. The *average* voltage, taken over each whole cycle, is, of course, nil when R is not a rectifier, because the negative half-cycle cancels out the positive. Using the perfect rectifier, the average of the positive half is 63 per cent. of the maximum, while the negative is, of course, nil ; so the average rectified DC taken over a whole cycle is 32 per cent. of the peak alternating voltage. Using the rectifier with the graph COB, the negative half-cycle (B in Fig. 77*b*) cancels out half of the positive half (*c*), which is two-thirds of the peak ; so the resulting D.C. average over a whole cycle is less than 11 per cent. of the peak A.C. One could present this information in the form shown in Fig. 78. Here the D.C. volts, obtained by averaging the output across the 1000 ohms, is plotted against the peak alternating voltage.

The characteristics of the rectifiers represented by FOJ (perfect) and COB are shown. The resistance characteristics—those shown in Fig. 76—of all real rectifiers are curved ; there is no abrupt change from low to high resistance. Look at Fig. 61. If the applied A.C. voltage

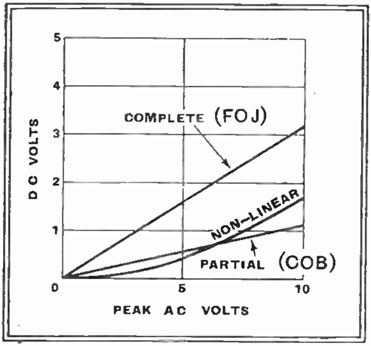


Fig. 78 : The average D.C. outputs given by applying A.C. to various rectifiers are shown in these characteristic curves, lettered to correspond with Figs. 76 and 77

is small, the rectifier is generally much less perfect than if it is large. So, instead of being straight lines, the graph in Fig. 78 is also curved : it might be like line COB to start with, rising nearly to the steeper slope of the perfect rectifier at large voltages. The effects of this will be considered in detail shortly. In the meantime we should remind ourselves that we are interested not in a single cycle, but in vast numbers of them per second ; so many that even over a very small space of time it is not possible to draw them individually. Fig. 79 shows them as a shading, and after a preliminary period in which they maintain a constant amplitude, representing an unmodulated carrier wave, they are modulated at an audio frequency (compare Fig. 75*a*). The D.C. output obtained from them by a perfect rectifier would be about one-third of the peak voltage, and is represented by a dotted line. Where the amplitude of the carrier fluctuates, the output fluctuates in proportion, being now an audio frequency voltage capable of operating phones or (if strong enough) a loudspeaker.

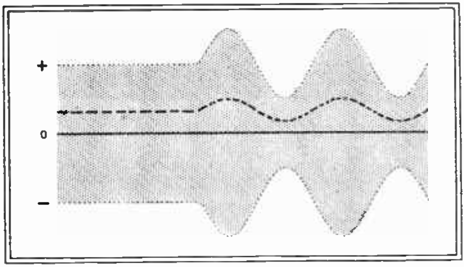


Fig. 79 : The dotted line indicates the rectified output obtained from the R.F. input, shown first unmodulated, and then modulated 50 per cent. The shading represents cycles at too high a frequency to be distinguished individually

83. The Simplest Complete Receiver

What is necessary, then, to form a complete radio receiver, is shown in Fig. 80. Here the modulated R.F. voltages, represented by the shading in Fig. 79, are developed across the tuned circuit; and the crystal rectifier, by virtue of its differing resistances to positive and negative half-cycles, passes through the phones a balance of current in one direction, which unidirectional current is at least approximately proportional to the amplitude from moment to moment of the waves radiated from the distant transmitter, and therefore is a copy of those set up by sound waves impinging on the microphone. And so the sound is reproduced.

In the simple crystal set of Fig. 80, the purpose of every component part, except the condenser C, has been indicated. In arriving at the steady or slowly varying voltage (or current) represented by the dotted line in Fig. 79 we have spoken of "averaging" the unequal positive and negative R.F. voltages resulting from the action of the rectifier.

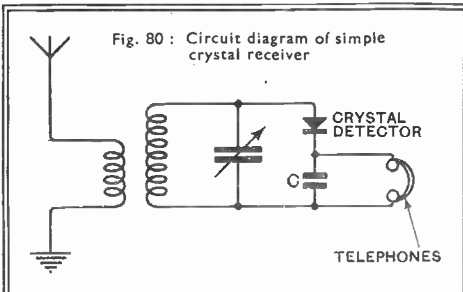


Fig. 80 : Circuit diagram of simple crystal receiver

The condenser does this averaging or smoothing out. When the applied voltage is near its peak in the forward direction the current through the phones cannot rise so rapidly (Sec. 19) and the spare current is taken in by the condenser which consequently increases its charge. During the intervals between one positive half-cycle and the next, when the current coming through the rectifier is nil or even negative, the condenser keeps things going by parting with some of its charge. It acts in much the same way as the bag in bagpipes, keeping them playing while the player draws his breath.

The circuit of Fig. 80 is that of a simple receiver, but it contains the kernel that every receiver must have. The two essentials are tuning, to select the required signal; and detection, to extract the A.F. currents from the received carrier. In addition, of course, phones or loudspeaker are needed to produce air waves from these currents.

We may add more tuned circuits to increase selectivity, and amplifiers operating on the signal either before or after detection, or both, to render the set more sensitive. But all these are mere elaborations; the crystal set described contains the essentials of tuning plus detection upon which every set, however ambitious, ultimately depends.

84. The Diode Detector

The thermionic diode valve (Sec. 63) has one strong point in its favour. Although it may not be perfect in the forward direction, for it always has some resistance, it is practically perfect in the reverse direction; that is to say, it passes negligible reverse current, even when quite large voltages are applied.

This fact enables the loss due to the forward resistance to be almost entirely avoided, and is bound up with choice of the resistance across which the output is derived—1,000 ohms in Fig. 77. In the simple set of Fig. 80, the aim is to obtain the maximum rectified *power*. The sound given by the phones depends on the strength of magnetic field produced in them, which in turn depends on the magnetising ampere-turns, which require both current and voltage (to force the current through a sufficient number of turns).

But in most sets, and practically all those employing valves, the rectified waves are not used directly in the sound-producing device but are amplified by a valve. Valves, as normally used, are voltage-operated devices; and therefore every effort is made to extract the maximum rectified voltage even if the current is thereby reduced.

Referring back to Fig. 77, and assuming the rectifier to have 500 ohms forward resistance and 2,000 ohms backward, a load resistance of 1000 ohms would have developed across it an average D.C. voltage equal to 10·6 per cent. of the A.C. peak, as already explained. Performing the same calculation for other values of load resistance, we get the result shown in Fig. 81, showing that with this imaginary rectifier a load resistance of 1,000 ohms gives the greatest voltage output, one-third of that for the perfect rectifier.

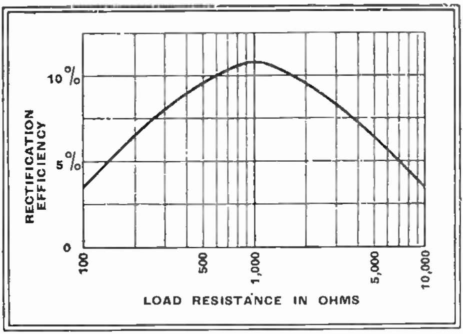


Fig. 81: Relationship between load resistance and rectification efficiency for the circuit Fig. 77c

If the backward resistance were infinitely large, then none of the forward rectified current would be neutralised by backward current; so even if the forward current were very small (due to a high forward resistance) the output voltage could, theoretically at least, be made to approach that given by a perfect rectifier, merely by choosing a sufficiently high load resistance. In practice, however, there are limits to the resistance that can be used. With the diode it is usual to make the load resistance 0·1 MΩ to 1 MΩ or even more, and the output approaches the theoretical maximum quite closely, provided that the applied voltage is not too small. In fact, by making use of the reservoir condenser (C in Fig. 80) it is possible to approach a rectification efficiency of 100 per cent., instead of the 32 per cent. that is the maximum without it; and so do nearly three times better than our "perfect" rectifier. This point is important enough to justify closer consideration.

85. Action of Reservoir Condenser

Suppose, in order to obtain the utmost rectified voltage, we made the load resistance infinitely great. The reservoir condenser would then be in series with the diode, with no resistor across it (Fig. 82). To simplify consideration we shall apply a square wave, as in Sec. 33, instead of a sine wave; and make its amplitude 100 volts. This applied voltage is shown as Fig. 83*a*. We shall also assume that its frequency is 500 kc/s, so that each *half* cycle occupies exactly one millionth of a second. At the start the condenser is uncharged, and therefore has no voltage across it. It can only acquire a charge by current flowing through the rectifier, which offers a certain amount of resistance, and so when the first positive half-cycle arrives its full 100 volts at first appears as a voltage across that resistance, as shown by the full line in Fig. 83*b*.

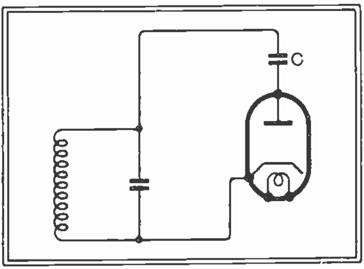


Fig. 82 : Diode detector with infinite load resistance

The current soon causes the voltage across the condenser to rise towards 100 volts, as shown by the dotted line. It is clear that the voltages across rectifier and condenser, being in series, must add up to give 100 volts so long as that is the applied voltage; and examination of Fig. 83*b* shows this to be so. The greater the resistance of the rectifier and the capacitance of the condenser, the longer the condenser takes to charge up, just as would a very large balloon inflated through a very narrow tube. As a matter of fact, if the capacitance in microfarads is multiplied by the resistance in megohms, the answer (known as the *time constant*) is the number of seconds required for the condenser voltage to reach 63 per cent. of the applied voltage. Suppose, then, that the rectifier resistance (assumed constant) is 0·01 MΩ and the capacitance is 0·0001 μF. Then the time constant is 0·000001 second, or one millionth of a second. In this case that happens to be the time occupied by one half-cycle of the 500 kc/s applied voltage. So at the end of the first positive half-cycle the voltage across the condenser is 63, while the voltage across the rectifier has dropped to 100 −63=37. Then comes the negative half-cycle. The diode ceases to conduct, and while the condenser therefore cannot charge up any more it likewise has no conducting path through which to discharge, and so remains at 63 volts until the second positive half-cycle. Meanwhile, the condenser voltage 63, together with the voltage across the rectifier, must be equal to the new applied voltage, − 100. The voltage across the rectifier must therefore be − 163.

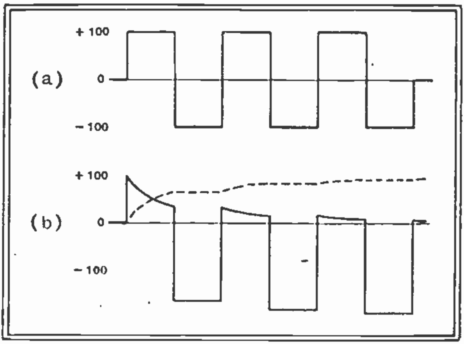


Fig. 83: Analysis of application of a square alternating voltage a to the circuit Fig. 82. The dotted line in *b* represents the voltage across C ; the full line, that across the diode

The net voltage applied to the condenser when the second positive half-cycle arrives is 100 − 63 = 37 volts, so at the end of this half-cycle the voltage across the condenser will increase by 63 per cent. of 37, or 23 volts, which, added to the 63 it already possessed, makes 86. The charge thus gradually approaches the peak signal voltage in successive cycles, while the average voltage across the rectifier falls by the same amount, as shown in Fig. 83*b*. The rectified output voltage is, therefore, 100 per cent. of the peak applied R.E. voltage. A similar result is obtained with a sine wave signal.

86. Choice of Component Values

This, of course, is excellent so far as an unmodulated carrier is concerned. When it is modulated, the amplitude alternately increases and decreases. The increases build up the condenser voltage still further ; but the decreases are powerless to reduce it, for the condenser has nothing to discharge through. In order to be able to follow the modulation it is necessary to provide such a path, which may be in parallel with either the condenser or the diode, so long as in the latter case there is a conducting path through the tuned circuit to complete the discharge route.

The resistance must be much higher than the diode forward resistance, or there will be loss of detector efficiency. Incidentally, there will also be damping of the tuned circuit. In practice one may reckon, as a close approximation, that the damping effect of a diode detector with a load resistance R is equivalent to connecting a resistance R/2 directly across the tuned circuit. It is for these two reasons that R is generally made not less than about 0·5 MΩ. But in order to follow modulation which may be as rapid as 10,000 c/s it is necessary for the time constant to be not greater than about one cycle of modulation, that is to say, 1/10,000th sec. The value of C is thereby fixed. If R is 0·5 MΩ, and C × R = 1/10,000, then C must be 0·0002 μF, which is in fact quite a likely figure. If the product CR is increased, the higher modulation frequencies cannot be followed and they are consequently reproduced less strongly than they should be. If it is reduced, then the detector efficiency falls off and the damping on the tuned circuit is increased.

Fig. 83 shows the process in slow motion, as it were. Speeding up the carrier frequency until it appears a mere blur, so that the variations in its amplitude due to modulation can be shown as in Fig. 84*a*, the charging up of the condenser to the RF peaks appears as at *b*, and the voltage across the rectifier itself as at *c*. Here the *average* is indicated by a heavy line and is the inverse of the curve *b*.

87. Varieties of Diode Detector Circuit

Two variations of the diode detector circuit have already been mentioned, and are shown in Fig. 85 *a* and *b*. The latter is necessary if the tuned circuit has a D.C. voltage between it and the cathode of the diode, as often happens in valve circuits (e.g., Fig. 101). As explained, the time constant of the condenser and resistor are such that the condenser voltage cannot follow the radio frequency to any great extent, but only the much slower audio frequency variations. Consequently most of the R.F. voltage delivered by the tuned circuit appears in the output along with the A.F. As will be seen later, this is a nuisance, and may be difficult to get rid of. So *c* and *d* are to be preferred where possible. Unfortunately *c* is not possible in those many cases in which it is essential for the tuning condenser

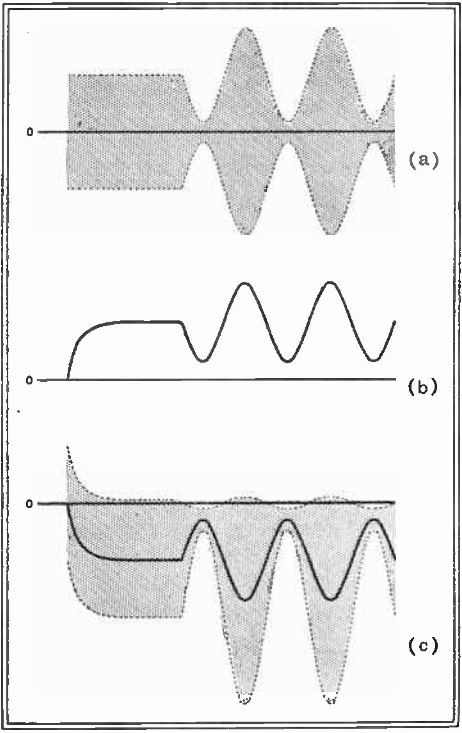


Fig. 84 : When a radio-frequency wave *a* (shown first unmodulated and then modulated) is applied to a diode detector having a time constant intermediate between the periods of the modulation frequency and the radio frequency, the voltage across condenser is as at *b*, and across diode as at *c*, where the full line is the average of the R.F.

to be earthed or at least at some steady potential. The circuit *d* avoids this difficulty, but runs into another, for the cathode is now at A.F. potential and the capacitance of the filament battery or heater transformer to earth may be objectionable, introducing hum. Sometimes, especially if followed by little amplification, *d* is satisfactory.

88. The Grid Detector

Mention of amplification brings us to the question of how to connect the diode detector to a stage of A.F.

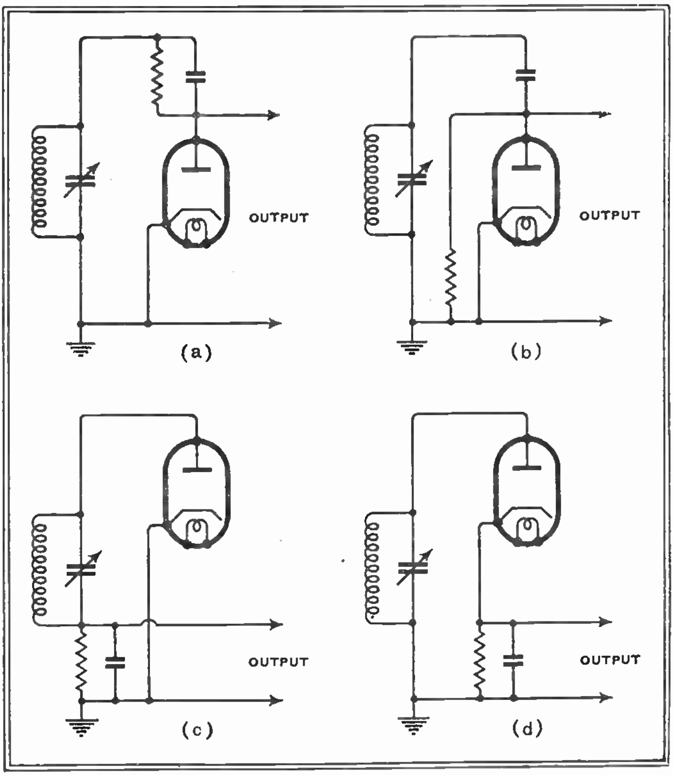


Fig. 85 : Various methods of connecting a diode detector to a tuned circuit

amplification (which may or may not be the valve that directly operates the loud speaker). When the detector circuit of Fig. 85*a* is connected straight to an amplifier, as shown in Fig. 86*a*, the anode-cathode path of the diode is in parallel with the grid-cathode path of the triode. Since both the anode of the diode and the grid of the triode consist simply of an electrode close to an emitting cathode, there would seem to be no need to have them both present. Experiment confirms this supposition; there is no change in the performance of the system if the diode is removed from its socket.

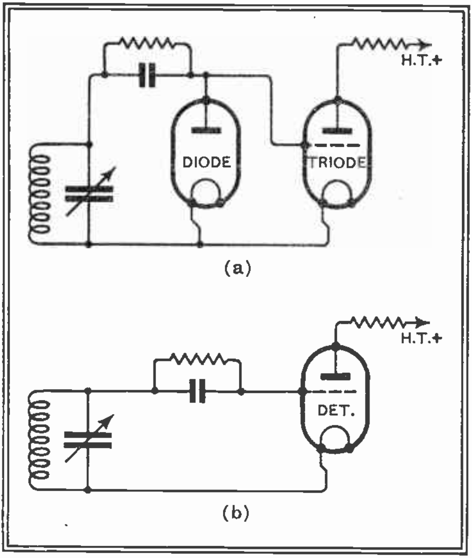


Fig. 86: Skeleton diagram of diode detector followed by triode as amplifier of the detected signals. Compare this with a "grid detector" (diagram *b*) in which cathode and grid behave as a diode detector, the valve then amplifying the detected signals

The simplified result is the well-known *grid detector* of Fig. 86*b*, in which the grid and cathode, acting as a diode, rectify just as in Fig. 85*a*. The A.F. voltages appearing at the grid as a result of this then control the electron stream through the triode and so produce an amplified voltage at the anode in the way described in Sec. 71.

In principle this is quite simple, but the two-fold function of the grid introduces a certain difficulty. If the triode were being used purely as an A.F. amplifier, the A.F.

input to the grid could be allowed to swing over the full range within which the characteristic is reasonably straight. For example, in Fig. 87, with anode voltage 60, it would be quite safe to swing from 0 to —2 volts, as shown by the heavy line, causing the anode current to vary proportionately between 4·8 and 1·2 mA. But we can see from Fig. 83*b* (and Fig. 84) that in order to cause an average voltage of —100 at the rectifier (the grid, in the case of a triode) it is necessary for the R.F. voltage to swing between 0 and —200. Similarly in Fig. 87 it is necessary for the R.F. to swing between 0 and — 4, shown shaded. Such a swing over-runs the anode current characteristic

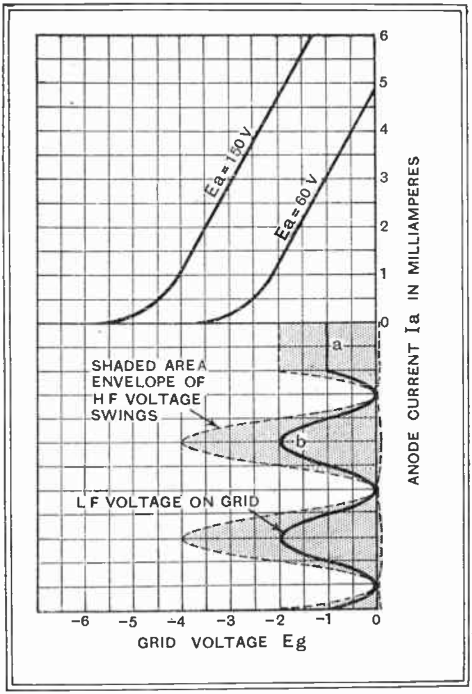


Fig. 87 : Curves showing reason for overload of grid detector. Even though the rectified voltage *b* on the grid may not overrun the straight part of the characteristic, the R.F. voltage accompanying it may cause anode-bend detection

and the negative peaks of grid voltage would fail to cause a *proportionate* reduction in anode current (for it cannot become less than zero). As the A.F. output is simply the average of the rectified R.F., its negative peaks also would be flattened off and distorted.

One answer is to keep the voltage of the applied R.F. signal down so low that it never drives the anode current, even momentarily, round the bend in the valve characteristic. This means that the A.F. output is restricted to one half of that which could be obtained if the R.F. were absent and the valve were acting purely as an A.F. amplifier.

89. Disadvantages of the Grid Detector

We shall see presently that restricting the input to the detector to not more than a volt or two tends to introduce another form of distortion, and also limits the usefulness of the detector for auxiliary duties (Chapter 17). To escape this dilemma, one method is to raise the anode voltage of the grid detector, as suggested by the 150-volt curve in Fig. 87. It can be seen that such an adjustment would accommodate the R.F. right up to the peaks of modulation, and so preserve the A.F. outline from distortion. This policy was at one time fairly popular under the name of *power grid detection*, but it may be noticed that to handle even such a modest A.F. as 1 volt peak the anode current runs up to an alarming figure, especially at times when there is little or no carrier wave.

The apparent economy of Fig. 86*b* has, therefore, proved illusory, especially as in many modern receivers it is desired to have a rectified voltage of 20 or more. So the diode has returned to favour, because there is no upper limit to the R.F. voltage that can be applied and satisfactorily rectified, and the R.F., when it has served its purpose, can be removed before the derived A.F. is passed on to the next stage.

90. Elimination of R.F.

There are two ways of doing this. One of them is to take the output from across the condenser as in Fig. 85*c* and *d*; because, as already pointed out, the condenser prevents the voltage across it from changing very much at such a high frequency as that of the carrier wave. The other, which enables Fig. 85 *a* or *b* to be employed, is to interpose a simple filter, such as that shown in Fig. 88, in which L is a R.F. choke coil and C a condenser so chosen that L has a much higher, and C a much lower, reactance to R.F. than to A.F. To rid the 85 *c* or *d* outputs of the last traces of R.F. such a filter is sometimes used with them too, but as the requirements are not so stringent it is common for L to be replaced by a resistor of 10,000 Ω to 50,000 Ω.

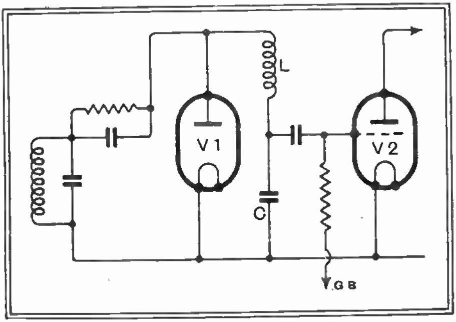


Fig. 88: Detection by diode V_1 followed by pure A.F. amplification by the triode V_2

91. Elimination of D.C.

If the explanation of Fig. 84*c* has been carefully followed it will be evident that in addition to the R.F., represented by the shading (which has now, we hope, been eradicated), and the A.F., represented by the waviness of the heavy line, there is also a negative D.C. component, represented by the lowering of the mean level of the heavy line below the zero mark by an amount approximately equal to the peak R.F. voltage. It is only the A.F. that we want to amplify ; so to cut out the D.C. a blocking condenser is connected between the diode and the grid of the amplifier, as shown in Fig. 88.

To enable this grid to be set to the most suitable bias for amplification (Sec. 73) it is connected to a source of bias voltage through a resistance, which must be kept high—generally about 1 MΩ— to avoid loading the detector unduly or introducing a particular form of distortion to be described shortly.

92. The Anode Bend Detector

As several causes of distortion have already been mentioned, it is about time to give some details of them. The

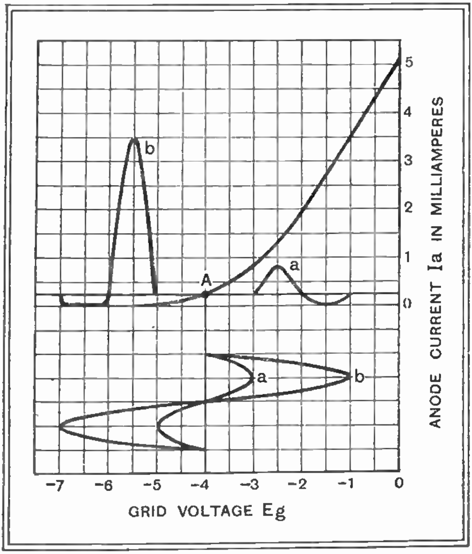


Fig. 89: Illustrating rectification of both weak and strong signals by an anode-bend detector

first can be illustrated by dealing with another type of detector. We have seen that the grid detector is really a diode rectifier, formed by grid and cathode working at about zero volts, where grid current starts, followed by a triode amplifier working on a straight (" linear ") part of its characteristic. An alternative is to bias the grid negative so that grid current never occurs, and to use the anode characteristic as the rectifier, which can be done by biasing it to a point at which anode current is reduced nearly to zero.

And here it is that we depart from the ideal and semi-ideal rectifiers of Fig. 76. Their characteristics, though they might be imperfect in not giving zero forward resistance and infinite backward resistance, at least change over abruptly from one to the other. But real rectifiers change more or less gradually, the characteristics having a curved portion as shown in Fig. 89 which represents a typical triode. For amplification, such a valve would be worked on the most linear part of the characteristic by means of biasing it about $-1\frac{1}{2}$ volts. The problem now is to find the most suitable bias for detection. Anode current appears to start at a bias of -5 volts, so this might be assumed to be a suitable working point.

But if the scale of anode current were expanded it would be found that a small current is flowing, so that if a small signal voltage were applied, swinging the grid between, say, $-5{\cdot}5$ and $-4{\cdot}5$, the "forward" current (grid going less negative) would be small and the "backward" current would not be zero; in other words the rectification efficiency would be low. A point such as A is more likely to be chosen, for although the backward current is appreciable its effect is outweighed by the much greater forward current.

It can easily be seen, however, that detection of very small signals is very poor indeed, for any small section of the characteristic curve shows little difference between the current changes produced by the positive and negative halves of the signal cycle. Even the signal marked *a* in the lower portion of Fig. 89, representing 1 volt amplitude in each direction on the grid, gives only a very partial rectification, as shown by the current curve *a*; whereas the 3-volt signal *b* gives a rectified current much more than 3 times as great. The rectified current, in other words, is not proportional to the amplitude of the carrier wave.

93. Effects of Curved Characteristics

The effects of this are two-fold: firstly, a receiver employing a detector of the nature described is insensitive to weak signals: secondly, the modulation of even fairly strong signals is distorted. In Fig. 78 we plotted the rectified output against the R.F. input voltage, and for the linear detectors whose characteristics are shown in Fig. 76 the results are also linear. But we have just seen that in the Fig. 89 type of detector, known for obvious reasons as the *anode bend* detector, the rectified output increases more rapidly than the applied R.F., and so we would get a rising curve.

The process is shown in greater detail in Fig. 90, which repeats the valve characteristic of Fig. 89, but instead of only a single cycle of R.F., the shading represents too many R.F. cycles to be shown individually, modulated at A.F. The unmodulated carrier wave, a short section of which is given at the top, has an amplitude of 2 volts, and swings the anode current (represented on the right) between 0 and 2 mA. around the initial current, which is 0·25 mA. The balance in favour of the positive half-cycles is, therefore, 1·5 mA. (peak values), and this represents the rectified output, and is indicated by a heavy line.

When modulated 25 per cent. the rectified output (positive swings less negative swings) gives a reasonably faithful reproduction of the modulation shape. But when

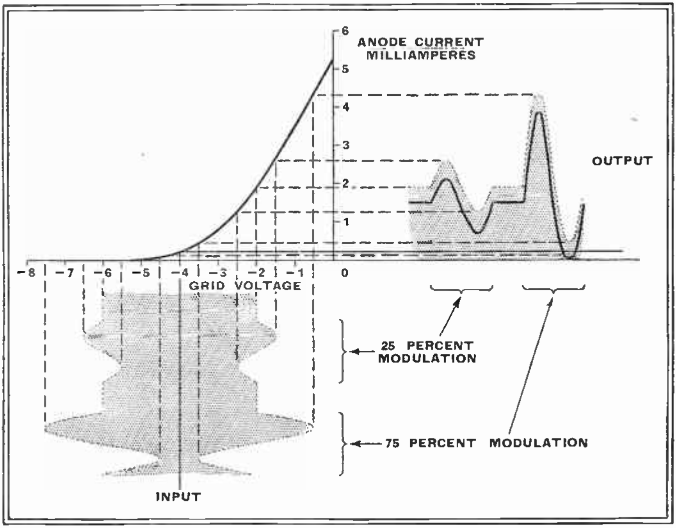


Fig. 90 : Extension of Fig. 89 to show application of R.F. carrier wave (modulated to differing depths). The full line represents the modulation-frequency rectified output, and can be seen to be not proportional to the original modulation of the input

the modulation depth is increased to 75 per cent. the rectified output increases considerably more for the positive peaks of modulation than it decreases for the negative. Note that at the latter the R.F. output is nearly equal in the positive and negative directions (above and below 0·25 mA) and the rectified output is, therefore, practically zero. Still greater modulation would, therefore, considerably increase the A.F. positive peaks, but could not reduce the negative which are already zero, so the distortion would be still more noticeable, and would result in falsification of the reproduced sound.

It is fairly easy to see that distortion occurs at a less modulation depth with small carrier wave amplitudes than with large ones. If an attempt is made to obtain distortionless reproduction up to fairly large modulation depths by increasing the carrier amplitude, it can be seen from Fig. 90 that there is a danger of running into grid current at the positive peaks of modulation, which would introduce still more serious distortion. Added to these drawbacks is another, described in Sec. 99 ; and so the anode bend detector is now seldom used except for special purposes.

94. A.C. Loading Distortion

The diode characteristics are not free from starting curvature, but are generally a considerable improvement on the anode bend controlled by grid voltage. Moreover, there is no restriction cn increasing the carrier amplitude,

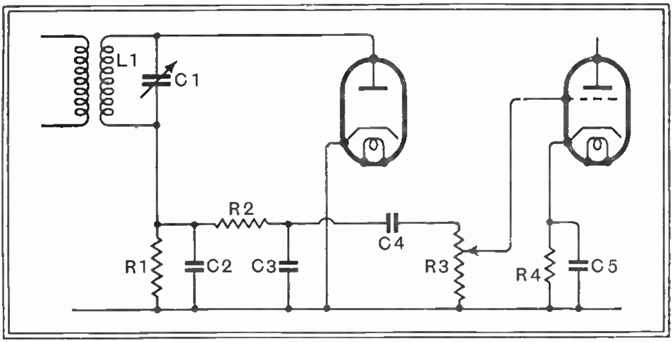


Fig. 91: Typical diode detector circuit followed by A.F. amplifier, showing that the A.C. load resistance is lower than the D.C., due to the path via C_4

and by doing so it is possible to reduce the distortion to negligible proportions even with 90 per cent. modulation (which is the deepest the transmitter itself is likely to be able to handle with tolerable distortion). As we have seen, this is not true of the grid detector, which causes distortion when the signal amplitude is either too small or too large.

The diode detector is, therefore, the most satisfactory, as correct operation of it is mainly a matter of arranging that the input is not too small. There is, however, one rather serious, but avoidable, source of distortion with the diode. Fig. 88 shows the blocking condenser which is used to prevent the D.C. component of the rectified output from reaching the amplifier valve. It is necessary to use a resistor on the far side of this condenser, for connecting to the appropriate grid bias source. This has also been found to be the most suitable point at which to introduce volume control, because distortion is kept down to a minimum by arranging for the signal to be as large as possible at the detector and as small as possible at the amplifier which follows it.

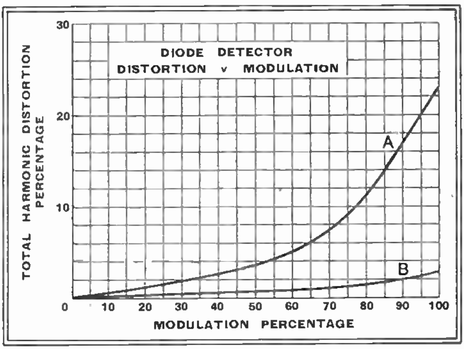


Fig. 92: Showing how the harmonic distortion in a detector circuit where the A.C. load resistance is two-thirds of the D.C. (curve A) is much greater than if A.C. and D.C. load resistances are equal (curve B)

So the most usual circuit is as shown in Fig. 91, in which L_1C_1 are the final R.F. tuned circuit, R_1 the diode load resistance and C_2 its condenser, R_2 and C_3 a R.F. filter, C_4 the blocking condenser, and R_3 the volume control. The purpose of R_4 and C_5 is to provide bias for the amplifier valve, as explained in Sec. 175. Note that in this circuit the D.C. load of the diode is R_1, whereas the A.C. load resistance is R_1 and $R_2 + R_3$ in parallel—the reactance of C_4 being assumed negligible in comparison. The A.C. load resistance is, therefore, lower than the D.C., and in these circumstances it is found that deeply modulated signals are distorted. The smaller the ratio of A.C. load resistance to D.C., the lower the percentage modulation at which distortion sets in.

Fig. 92 (due to Langford Smith's "Radio Designer's Handbook") shows the great increase in distortion resulting from making the A.C. resistance one-third less than the D.C.; as, for example, making R_1 in Fig. 91 0·5 MΩ and $R_2 + R_3$ 1 MΩ. This is, therefore, a point that must be carefully watched in the design, especially as other A.C. loads may be connected in parallel (see Chapter 17).

CHAPTER 10

THE SINGLE-VALVE SET: REACTION

95. The Circuit

BY now we have covered enough ground to be able to discuss the behaviour of a simple type of receiver. This will take us away, for the first time, from the fairway of simple theory, and we shall find ourselves making acquaintance with some of the incidental complications that arise when we have to deal with real circuits in place of circuits idealized to bring out their fundamental properties.

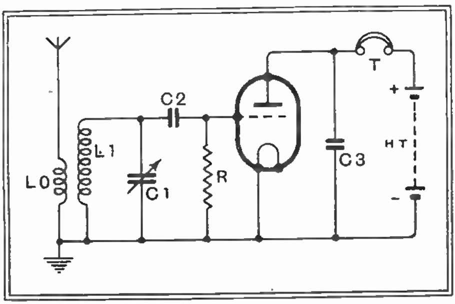


Fig. 93: Circuit of single-valve set in which a triode valve is used as grid detector and amplifier of the detected signals

Fig. 93 shows the circuit of a single-valve receiver. The outline of its working is simple enough. The currents induced in the aerial by the received wave flow through the primary winding L_0, to which is coupled the secondary winding L_1, this being tuned to the frequency of the desired signal by adjustment of the variable condenser C_1. The signal-voltage developed across the tuned circuit is applied, through the grid-condenser C_2, between grid and cathode of the triode valve which, since the resistance R (the *gridleak*) is returned to cathode, will behave as grid detector. The detected and amplified signals in the anode circuit are passed through the telephones T and so made audible to the listener.

The function of C_3 is the subject of Sec. 100. The letters "HT" against the anode supply battery stand for "high tension," which in the early days of valve technique was the name given to this battery to distinguish it from the filament battery or "LT." Although the name is undesirable, because in electrical engineering its use is definitely allocated to much higher voltages than those in receiving sets, it has gained too strong a hold to be quickly abolished.

96. The Radio-Frequency Transformer

In a radio-frequency transformer, such as is made up by L_0 and L_1 in Fig. 93, the ratio of turns on the two windings has to be adjusted to suit the needs of the circuit in which the transformer is to be used. The secondary, being tuned, has to have the right number of turns to give it the inductance necessary to cover the wave-band over which it is desired to tune ; that leaves us with the primary as sole variable.

In Fig. 94 *a* we have a tuned circuit of dynamic resistance R, shown as a resistanceless coil and condenser shunted by R as a load-resistance. This coil contains n times as many turns as the primary, and the two are supposed to be closely coupled. The impedance of the primary as such, since it is untuned, is in all practical cases minute compared with the transferred load from the secondary ; we shall therefore be safe in assuming that the primary is equivalent to a resistance of R/n^2 ohms. (Sec. 48.) Similarly, if we put a resistance R_0 across the primary, the secondary will behave as though a resistance n^2R_0 had been put in parallel with it. A smaller value of n, which means a larger primary, therefore increases the damping introduced into the secondary by the resistance of an aerial or of a valve connected to L_0.

Where the source of voltage V_0 has no resistance, as in Fig. 94 *a*, the voltage across R will rise steadily as the ratio n is increased, the rise in power corresponding to this being made up by a larger current in L_0. But if, as in all practical cases, the generator has an internal resistance, as at *b* in the figure, this rising current will reduce the voltage at the terminals of L_0, on account of the voltage lost across R_0.

It can be shown that maximum voltage is developed across R when the ratio of the turns is so chosen as to make the effective resistance of the primary equal to R_0, so that exactly half of the total generator-voltage appears on L_0. If, for example, L_0 were connected in the anode circuit of a valve of anode resistance 20,000 ohms, and the dynamic resistance of the tuned secondary were 180,000 ohms, the value of n for greatest amplification would have to be such as to make $\frac{180{,}000}{n^2} = 20{,}000$, to do which n must clearly be 3. If L_1 had 90 turns, there would therefore have to be 30 turns on L_0. Fig. 94 *b* shows the distribution of voltage with n chosen for maximum voltage across R.

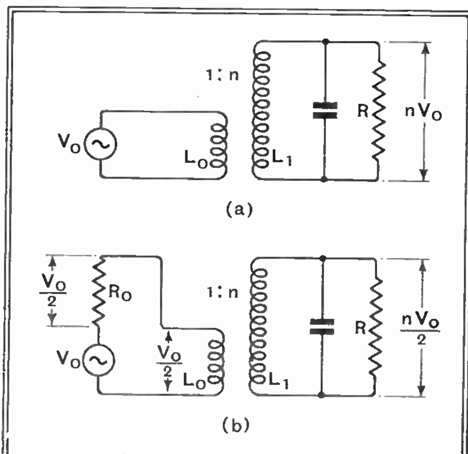


Fig. 94: (*a*) R.F. transformer. The tuned secondary, of dynamic resistance R, is shown as a loss-free circuit shunted by R as load-resistance. (*b*) Where the source (aerial, valve, etc.) has resistance, maximum volts on R are obtained when *n* is so chosen as to make the apparent resistance of the primary equal to that of the source. For this, $n = \sqrt{\frac{R}{R_o}}$, and the voltage distribution is as shown

The same relationship works backwards too; putting a valve of

20,000 ohms in parallel with L_0 is equivalent to putting a second resistance of 20,000 $\times$ n^2(= 180,000 ohms) across the tuned circuit, thereby halving its dynamic resistance and correspondingly flattening its tuning.

97. Effect of Primary Turns

Applying this to the aerial transformer of Fig. 93, though perfectly easy in theory, is of little use in practice, because the aerial is not a simple resistance, but has characteristics which change over quite a wide range as we vary the wavelength by tuning. Moreover, the theory of Sec. 96 assumes that *all* the magnetic lines of force due to the primary link with the secondary ("100 per cent. coupling"), whereas in practice and especially with an air-core RF transformer this is never quite so. But the theory does very clearly show that for each wavelength there is a definite number of turns for L_0 which will give the greatest voltage on L_1. In radio receivers, maximum signal voltage is not everything; it may be, and often is, worth while to sacrifice some in favour of selectivity. Increasing the turns above the optimum number will give less voltage, and at the same time will transfer a greater proportion of the aerial resistance into the tuned circuit, so making it tune more flatly. In such a case we say that the aerial is *too closely coupled* to the tuned circuit.

Reducing the turns below the optimum will again reduce the voltage on L_1, but the aerial will now be rather *loosely coupled* to the tuned circuit, so that the damping passed on to it will be but small. Since at least one gets some extra selectivity in exchange for loss of volts when one couples loosely, while over-tight coupling loses us both volts and selectivity, it is usual to keep the turns on L_0 down to a fairly small number. In practice the matter is generally still further complicated by the need for avoiding appreciable alteration in tuning due to the connecting of any aerial likely to be used, in order to enable the tuning condenser to be ganged (mounted on the same shaft) with condensers tuning other circuits in the receiver. So the designer has to make the best of a rather complex problem.

98. Tuning Range

Reverting to the circuit of Fig. 93, we see that the total capacitance across L_1 is increased above that of the tuning condenser C_1 by the extra capacitances due to the valve and its holder, the wiring, the terminals or tags to which the ends of L_1 are brought, and by a certain amount of capacitance transferred from the aerial through the primary L_0. If the maximum capacitance of C_1 is the usual 500 $\mu\mu$F, the total will be about 550 $\mu\mu$F, from which we find that if we are to tune up to 550 metres L_1 must have an inductance of about 155 μH.

The lowest wavelength to which a circuit will tune is almost entirely a function of the extra, or "stray" capacitances, but in any average case a coil of 155 μH. will just comfortably tune down to 200 metres.

As we have seen (Sec. 86), usual values for C_2 and R are 0·0002 μF and 0·5 MΩ. The effect of the grid circuit of the valve in damping the tuned circuit has also been mentioned; it is approximately equal to putting across C_1 a resistance $\frac{1}{2}$R, or, in this case, 0·25 MΩ. If the dynamic resistance of the tuned circuit alone is 125,000 ohms it will be reduced by this damping to two-thirds of its original value. Alternatively expressed, the valve will increase the equivalent series resistance of the circuit by 50 per cent.

99. The Miller Effect

In addition to this effect, which is solely due to the grid current taken by the valve, there is another which depends for its existence upon the voltages developed in the anode circuit, and upon the small capacitance between the anode and the grid of the valve. In Fig. 95 *a* there is shown the conventional diagram of the valve used as radio frequency amplifier, the impedance in the anode circuit being represented by Z_a. This may be a resistance, a capacitance, or an inductance. C_{ga} represents the total capacitance between grid and anode, which is partly in the valve-electrodes themselves and the glass pinch supporting them, and partly in the valve-base, the valve-holder, and the wiring.

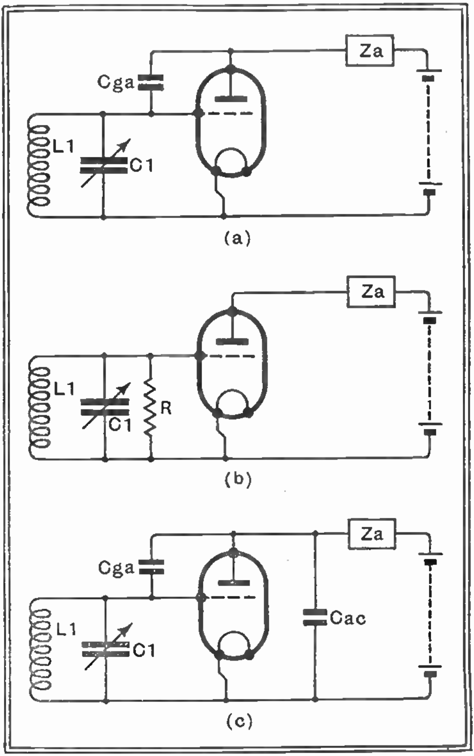


Fig. 95 : Illustrating the Miller Effect. *a* A voltage at the anode of a valve can pass, by way of C_{ga}, back into the grid circuit. *b* If Z_a is a capacitance, the anode-grid feed is equivalent to connecting a damping resistance R across the grid circuit. *c* Owing to the stray capacitances from anode to cathode, any anode-circuit impedance Z_a is necessarily shunted by C_{ac}

Since the amplifying action of the valve produces a radio-frequency voltage at the anode, a small radio-frequency current will flow through C_{ga} and the tuned circuit to the cathode of the valve. In flowing through the components in the grid circuit, this current will develop across them a voltage, and this voltage might have any one of three possible phase-relationships with the voltage already present due to the signal. If it were in phase with the signal voltage the two would simply add, and the original voltage would be artificially increased. If it were 180° out of phase, on the other hand, this new voltage would be in opposition to that already there, and the energy fed through C_{ga} would tend to damp out and reduce the signal voltage. In the third case, which is of less interest, the voltage fed back from the anode is 90° out of phase with that already present, and would therefore neither help nor hinder it.

In general there will be a combination of this with one of the other two.

Since the alternating anode current of the valve is produced in response to the alternating voltage at its grid, the phase of the current is fixed with respect to the original signal voltage. The phase of the alternating voltage developed on the anode depends on the nature of the impedance Z_a, through which the current is made to flow. It can be shown that if Z_a is a pure resistance, and so long as the grid circuit is tuned exactly to resonance, the phase of the current fed back through C_{ga} is such as to increase the apparent capacitance of the grid. If $A\left(=\frac{\mu R}{r_a + R}\right)$ is the stage gain, then for one volt applied to the grid $-A$ volts appear at the anode. The *change* in voltage between grid and anode is therefore not 1 volt, as it would be if the anode were connected straight to the battery, but $(1 + A)$ volts. The quantity of electricity required to charge the condenser formed by grid and anode is therefore $(1 + A)$ times as great, which is equivalent to multiplying its actual capacitance by $(1 + A)$. If a large amplification is attempted, this effect becomes very serious.

If Z_a is a capacitance, the energy fed back tends to damp out the voltage already present, while if it is an inductance, the energy fed back reinforces and increases the signal voltage on the grid.

In the case where Z_a is a capacitance, the damping effect on the grid circuit can be exactly reproduced by connecting a resistance R of suitably chosen value across grid and cathode of the valve in the manner shown in Fig. 95 *b*. But a little thought will make it clear that since the whole effect depends on the alternating voltage at the anode, changes in magnitude of this will alter the value of the equivalent damping resistance R. The higher the impedance of Z_a (or since we are considering the case where this is a capacitance, the lower the value of this capacitance) the higher will be the voltage developed, and hence the greater will be the damping effect in the grid circuit. Thus a high value of Z_a corresponds to a low value of R, reducing very markedly the voltage across the tuned circuit of Fig. 93, and flattening its tuning to a considerable extent.

If the capacitive reactance of Z_a is made high (corresponding to a small value for C_3 in Fig. 93) the damping can be very serious indeed ; with C_3 omitted altogether, so that the anode-circuit impedance for radio-frequency currents consists only of the stray capacitances across valve, valve-holder, and telephones, the energy fed back from anode to grid may be equivalent, for a signal at 1,000 kc/s, to connecting a resistance of as low a value as 5,000 Ω between grid and cathode. Since the dynamic resistance of the tuned circuit L_1C_1 will probably be twenty times as great as this, the effect of the damping in dropping signal strength and flattening tuning is positively catastrophic.

100. The Anode By-pass

This explains the presence of C_3 in Fig. 93 ; it is inserted as a low impedance to the radio-frequency currents so that the voltage developed at the anode may be as low as possible.

Evidently there is a limit to the reduction in damping that can be effected in this way, because although from the radio frequency point of view the higher the capacitance of C_3 the less the voltage developed across it and the less the damping thrown back into the grid circuit, we have to remember that there are audio frequency voltages that we want. Fortunately the reactance of C_3 is much greater at audio frequencies, so they are not suppressed so effectively as the radio frequencies. If we make C_3, about 0·001 μF, its reactance at 1,000 kc/s will be little more than 150 ohms, while at the higher audio-frequencies (5,000 cycles) it will rise to 30,000 ohms, which will not be a very serious shunt to the telephones, and so will not cause too great a diversion of high notes from their windings. Like almost every other point in a wireless set, the choice of a capacitance for C_3 is a compromise that tries to make the best of both worlds.

101. Reaction

Instead of striving to prevent feed-back from the anode to the grid circuit, we can deliberately introduce it, so arranging matters that we have it at all times completely under control. This can be done by inserting in the anode circuit a coil L_2, as in Fig. 96. As the diagram shows, this coil is close to L_2, in order to couple with it inductively, while the arrow running through them indicates that their relative positions can be adjusted as required.

Part of the radio-frequency current flowing in the anode circuit will pass direct to cathode through C_3, and part will will flow through L_2, the capacitance C_4 across the phones, and the anode battery. This latter portion, in its passage through the coil, sets up round L_2 a radio-frequency field which, in passing also through L_1, induces a voltage in the latter. By connecting L_2 in the right direction this voltage can be made either to assist or to oppose the voltage there already, the effect in either case becoming more marked as L_2 is brought closer to L_1. We will consider some of the effects that arise when the feed-back assists the original voltage in the grid-circuit.

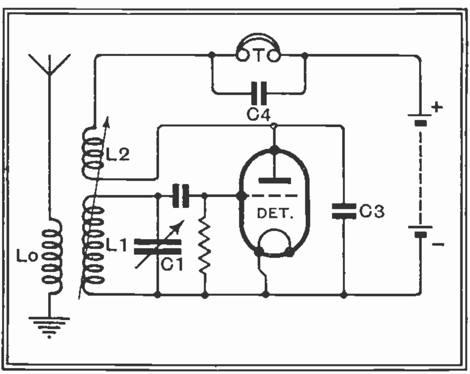


Fig. 96 : Conventional single-valve set with adjustable reaction

Since the amount of energy fed back can be controlled by adjusting the coupling between the two coils, let us begin by supposing that this adjustment has been made in such a way that the signal-voltage across L_1 has the same value, no matter whether the valve is connected to it or not. This implies that the reduction of voltage occasioned by grid-current damping and by the energy fed back through the anode-grid capacitance is exactly offset by the voltage fed back from L_2.

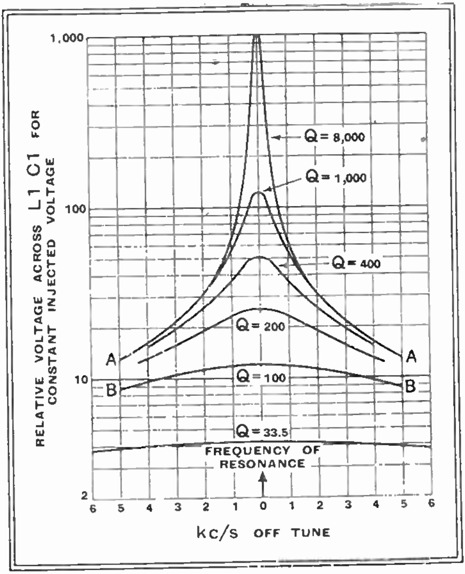


Fig. 97 : Showing relative voltage at resonance (height of peak) developed across L_1C_1 for various coil magnifications Q

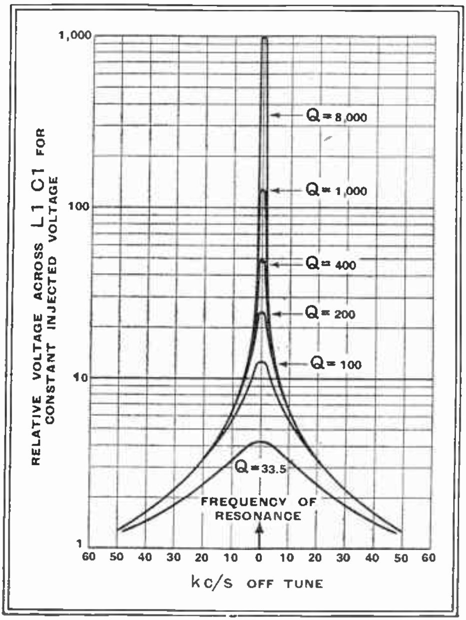
We have already seen (Sec. 86) that the damping due to the valve can be exactly imitated, both in its effect in reducing the voltage across L_1 and in its effect of flattening the resonance curve of the tuned circuit, by connecting a resistance across the tuning condenser. From our knowledge of radio-frequency resistance, we are aware that the effect of any parallel resistance can be duplicated by opening the tuned circuit (between L_1 and C_1) and inserting a series resistance of equivalent value (Sec. 61). And now we see that the valve-damping can be neutralized again by a suitable coupling between L_2 and L_1.

We conclude that by feeding into it energy from the tuned circuit of a valve it is possible to *neutralize resistance* in a tuned circuit connected to its grid. This neutralization of resistance is in this country called *reaction* (to be sharply distinguished from *reactance !*) or sometimes *retro-action*, and in America is called " regeneration ".

Of course, reaction does not neutralize resistance in any strictly literal physical sense. The sole characteristic of resistance is its absorption of power ; if, therefore, we supply power from the anode circuit of a valve the circuit in which that resistance is located behaves as though it had lost some of its resistance. The valve is used as a source because it is only by making the voltage itself (in the grid circuit) control the power used to enhance it that the two can be locked unalterably together in the required phase.

In discussing tuned circuits (Sec. 54) we saw that reduction of radio-frequency resistance increases both the magnification and the selectivity of a tuned circuit. With the aid of a valve to provide reaction we are now in a position to adjust the resistance of the tuned circuit L_1C_1 to any value that takes our fancy, simply by approaching L_2 cautiously towards L_1 until the resistance has been reduced to the desired extent. As we do this the voltage de-

Fig. 98 : Extension of Fig. 97, showing the voltages across L_1C_1 when tuned exactly (peak) and when detuned to various extents



veloped by the signal across L_1C_1 will steadily rise and the tuning will become steadily sharper.

The effect on the tuned circuit can best be visualized with the aid of a series of resonance curves. In Figs. 97 and 98 the voltage across L_1C_1 is plotted against frequency for a number of values of magnification. A glance will show that as the magnification is increased by the application of reaction the signal-voltage rapidly rises* and the sharpness of tuning, as measured by the ratio of the voltage at resonance to that developed a few kilocycles off tune, becomes greater.

So great is the increase that it could not be shown on a curve sheet like those in Chapter 6: a logarithmic scale of relative voltage is necessary, and compared with Fig. 53 the rise is greater than it looks.

The difference between the two sets of curves in Figs. 97 and 98 is purely one of frequency scale; in the former the frequency scale extends only to 6 kc/s on either side of resonance, so that only the peaks of the curves are plotted. In the latter the behaviour of the circuits is shown over a range of 60 kc/s each way from resonance. In both cases the lowest curve is a fair representation of the behaviour of a tuned circuit of normal radio-frequency resistance connected to a detector. With the reaction coil out of use the circuit assumed has $L = 155\mu H$, $r = 10\ \Omega$, and is supposed to be tuned to 1,000 kc/s. Detector damping across it is taken as 50,000 Ω. For the tuned circuit alone $Q = 98$, $R = 95{,}000\ \Omega$; with detector damping in parallel the total dynamic resistance is reduced to 32,600 Ω, making the equivalent R.F. resistance 29·2 Ω and reducing the effective magnification to 33·5. The curve next in order ($Q = 100$) represents the same tuned circuit with the effects of detector damping almost exactly offset by the judicious application of reaction. Successive curves show the effect of more and more reaction, culminating in the extreme case where the magnification has been increased to 8,000, which is about

* The relative heights of the peaks are calculated on the basis of constant injected voltage. This ignores the reaction of L_1C_1 upon L_0, the aerial primary.

the highest value known to have been reached, and held, by this means. It corresponds to the neutralization of all natural resistance of the circuit except for a small residue of about one-eighth of an ohm.

102. Over-Sharp Tuning

At first sight it would appear that the reduction of circuit resistance, even to such very low limits as this, was all to the good, since it would increase both the sensitivity and the selectivity of the receiver. If we had to receive a simple carrier wave this conclusion would be true, but we must remember that the signal from a broadcasting station consists of a *modulated* carrier. As we have seen, the modulation consists in a variation in the amplitude of the carrier at the frequency of the musical note it is desired to transmit. We know also (Sec. 30) that if a tuned circuit had no resistance at all, any oscillation that might be set up in it would persist after the removal of the signal source, unchanged in amplitude, for ever. Such a circuit would evidently be quite incapable of following the rapid variations in amplitude of a modulated carrier.

It follows, therefore, that as we approach towards zero resistance by a greater and greater application of reaction, the voltage across the tuned circuit will tend more and more to " hang ", following with greater and greater sluggishness the variations due to the modulation. For the highest audible notes the radio-frequency voltage has to change in amplitude most rapidly ; as the resistance of the tuned circuit is decreased these will therefore become weak and vanish at a value of resistance still high enough to enable the low notes, for which the variations in amplitude of the carrier are proportionately slower, to remain substantially unaffected.

The high, sharp peak of a very low-resistance circuit such as that giving the curves " $Q = 8{,}000$ ", therefore, tells us that high modulation-frequencies cannot be followed. On the other hand, the flatter curves such as that for $Q = 100$, indicate a resistance high enough for any current through the circuit to die away rapidly unless maintained by a driving voltage, thus enabling the voltage-variations across L_1C_1 to be a faithful copy of the signal as received from the aerial.

103. The Theory of Sidebands

By regarding the modulated wave from a slightly different point of view, the relationship between sharpness of tuning and the loss of high audible notes can be shown to be very much more intimate than has been suggested. Strictly speaking, it is only an exactly recurrent phenomenon that can be said to possess a definite frequency. The continuous change in amplitude of the carrier wave in response to modulation makes the radio-frequency cycle of the modulated wave non-recurrent, so that in acquiring its amplitude variations it has lost its constancy of frequency.

A mathematical analysis shows that if a carrier of f_1 cycles per second is modulated at a frequency f_2 cycles per second the resulting modulated wave is exactly equivalent to three separate waves of frequencies f_1, $(f_1 - f_2)$, and $(f_1 + f_2)$. It is not easy to perform the analysis of the modulated wave into its three components by a graphical process, but the corresponding synthesis, adding together three separate waves, requires nothing more than rather extensive patience.

Fig. 99 shows at *a*, *b*, and *c* three separate sine-waves, there being 25, 30, and 35 complete cycles, respectively, in the length of the diagram. By adding the heights of these curves point by point, the composite curve at *d* is obtained. There are in its length 30 peaks of varying amplitude, and the amplitude rises and falls five times in the period of time represented on the figure. If this is a thousandth part of a second, curve *d* represents what we have come to know as a 30 kc/s carrier modulated at 5,000 cycles.

Thus a carrier modulated at a single audio-frequency is equivalent to three simultaneous signals, the un-modulated carrier itself and two associated steady frequencies spaced away from the carrier on either side by the frequency of modulation. In the case of a musical programme, in which a number of modulation frequencies are simultaneously present, the carrier is surrounded by a whole family of extra frequencies. Those representing

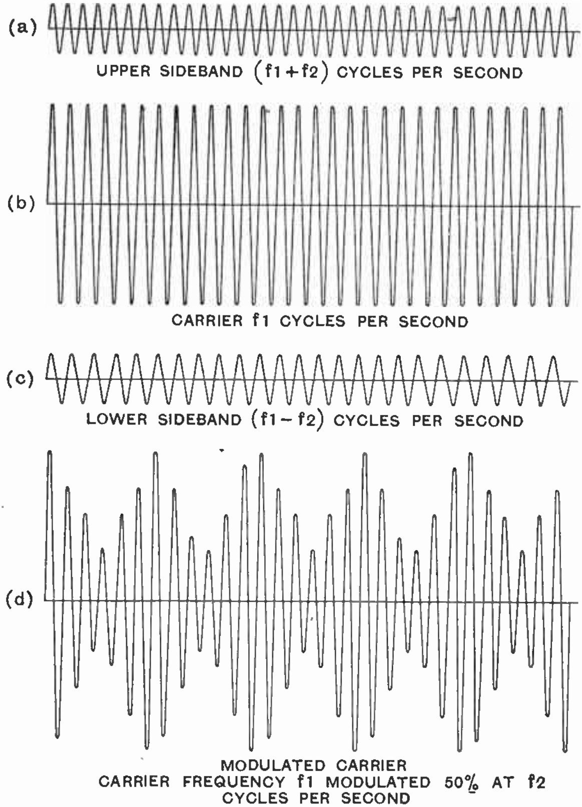


Fig. 99 : Showing the relationship of a modulated carrier *d* to its three components

the lowest musical notes are close to the carrier on either side, those bringing the middle notes are further out, and the highest notes are the farthest removed from the carrier frequency. The spectrum of associated frequencies on either side of the carrier is called a *sideband*, and as a result of the presence of these a musical programme, nominally transmitted on a (carrier) frequency of 1,000 kc/s, will spread over a band of frequencies extending from about 993 to 1,007 kc/s.

We now have a direct relationship between the selectivity of a tuned circuit and its ability to receive the highest notes likely to be present as modulation on the carrier. If the resonance curve of the circuit is not substantially flat over a central portion wide enough to include the whole of the required sidebands, high notes will be attenuated—they will be quite literally tuned out owing to over-selectivity. In the curve for $Q = 8{,}000$, in Fig. 97, the sidebands corresponding to a modulation frequency of 5,000 cycles are shown, at points AA, as being transmitted at about 1·3 per cent. of the central carrier frequency. Lower notes are more fully transmitted, higher notes even more greatly attentuated. The result will be "woolly" and more or less unintelligible speech, and "boomy" music. For a tuned circuit in which $Q = 100$, however, 5,000-cycle notes are passed at 70 per cent. of the carrier amplitude (BB in Fig. 97).

It is clear from these considerations that high selectivity is not altogether an unmixed blessing in the reception of telephony, and that too great an application of reaction will sharpen tuning to such a point that the quality of the received programme suffers badly. Nevertheless it remains invaluable for neutralizing the losses due to detector damping, and may, without serious detriment to quality, be pressed far enough to halve or even quarter the natural resistance of a tuned circuit. But much greater amplification than this is needed for the successful reception of distant transmitters.

CHAPTER 11

RADIO-FREQUENCY AMPLIFICATION: SCREENED VALVES

104. Increasing Range

IF we want to increase the range of our single-valve set sufficiently to enable us to receive transmissions from distant stations, the only alternative to reaction is amplification by a valve. A valve may be used in either of two ways: it may be applied to amplify the modulated radio-frequency signal before detection (radio-frequency amplification) or it may be made to amplify the detected audio-frequency signal (audio-frequency amplification). The choice between these two alternative methods is dictated by the characteristics of the detector.

We know that a large signal can be detected with less distortion than a small one; it is also true that any detector is very insensitive to really weak signals (Secs. 93 and 94). Unamplified signals from a distant station (a millivolt or less) would swing the grid of a detector over a portion of its curve so small that it would be virtually a straight line over that tiny range. We are driven, therefore, to amplify weak signals before detection in order to provide sufficient input to operate the detector satisfactorily.

105. Simple Resistance Coupling

At first sight it might seem that, since a resistance behaves alike to currents of all frequencies, one would obtain very satisfactory results by coupling valves together for radio-frequency amplification in the manner suggested

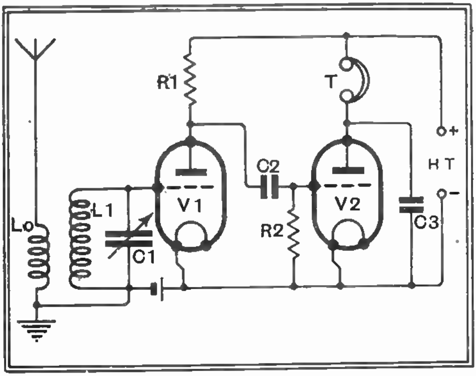


Fig. 100: The resistance-coupled R.F. amplifier is of little more than academic interest

in Fig. 100. Here is shown a stage of resistance-coupled amplification preceding the detector valve V_2. The valve V_1, receiving its signal from the secondary of the aerial transformer L_0L_1, produces an amplified voltage across the resistance R_1 in its anode circuit. This voltage is conveyed to the grid of V_2 through the condenser C_2 which, while readily passing R.F. currents, protects the grid of V_2 from the steady positive voltage at the anode of V_1.

If one could build this receiver without departing from the strict letter of the circuit diagram it would work very well. Unfortunately, there appear in a practical set the stray capacitances from anode to cathode of V_1, and from grid to cathode of V_2, together with wiring, etc. They make, in an average case, a total of 40 $\mu\mu$F or more, which provides at 1,000 kc/s a path of reactance about 4,000 Ω. This sets an upper limit, irrespective of the value adopted for R_1, to the anode-circuit load of V_1. With so low a load V_1 will not provide very high amplification ; one may expect a gain of about five times with an average valve.

But even this is not the worst fault of the circuit of Fig. 100. The anode circuit of V_1 being predominantly capacitive, it damps the tuned circuit L_1C rather heavily. In the case of the detector, we reduced this damping to reasonable limits by inserting a condenser direct from anode to earth in an attempt to reduce the radio-frequency voltage at the anode as nearly as possible to zero ; we were then wanting only the rectified audio-frequency signals. In the present case we obviously cannot do this, or we shall short out the signals, and in consequence of the development of an appreciable radio-frequency voltage at its anode, V_1 is equivalent to a damping resistance of the order of 6,000 Ω across the tuned circuit. If the initial dynamic resistance of this, undamped, were 120,000 Ω, the introduction of this damping would reduce the voltage across it to less than one-twentieth.

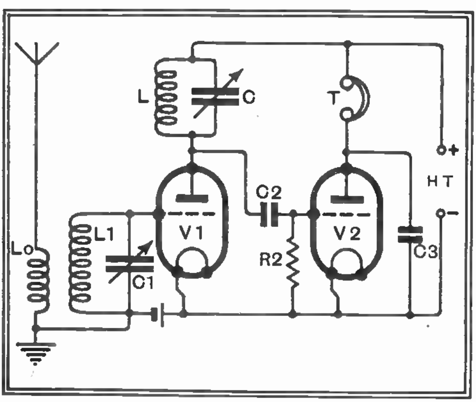
With V_1 amplifying this reduced signal five times, the voltage finally delivered to V_2 would be one-quarter of that developed across L_1C_1 unloaded. On the whole, not a very successful amplifier.

The replacement of R_1 by a radio-frequency choke, making a choke-coupled amplifier, leaves the problem untouched ; the faults of the circuit lie in the stray capacitances across the anode load and in the anode-grid capacitance of the valve, and not in the type of coupling used.

106. The Tuned Anode Circuit

But if we can find a method of neutralizing the effects of stray capacitances we shall be in a better position. Such a method lies ready to hand; we have only to place in parallel with them (i.e., from anode to earth or to the H.T. line) a coil of reactance equal to that of the stray capacitance, thereby forming a tuned rejector circuit. To

Fig. 101 : Tuned anode R.F. coupling. Compare with Fig. 100 and note that the various stray capacitances are now in parallel with C and so form part of the tuning capacitance



avoid the awkwardness of having to readjust the value of this inductance every time we want to tune from one wave-length to another, we add a variable condenser for tuning. This gives us the *tuned anode* circuit of Fig. 101.

The diagram shows that LC is connected, as a complete circuit, between the anode of the valve and its battery. The stray capacitance in parallel with this now has no more effect than to make it necessary to reduce C itself a little below the value at which tuning would be attained in the absence of the strays. The whole forms a simple parallel tuned circuit. For the frequency of resonance we have seen that this behaves as a pure resistance R, the dynamic resistance L/Cr of the circuit. We have therefore worked our way back, so far as the electrical behaviour of the system is concerned, to the unrealizable resistance-coupled arrangement of Fig. 100. The amplification given by a tuned anode stage will be that calculated from the simple formula $A = \frac{\mu R}{R + r_a}$ given in Sec. 72 for a resistance-coupled stage, but we must now interpret R as the dynamic resistance of the tuned circuit.

We have found a remedy for the effects of stray capacitance in limiting amplification, for the circuit of Fig. 101 will give a gain of some 25 or 60 times with battery or mains valves respectively, even if R is no more than 100,000 ohms. It remains to be seen whether the anode-grid capacitance is equally harmless.

107. Grid-Anode Capacitance

So long as the anode circuit is exactly tuned to the frequency of the signal being received, the anode circuit of the valve will be purely resistive, and voltage fed back through C_{ga} (Fig. 102) will neither assist nor damp down the voltage on the grid but will make it necessary to use less of C_1 in order to compensate for Miller effect. If the applied frequency (or alternatively the capacitance of C) is now increased, slightly more current will flow through C than through L, so that the anode circuit becomes capacitive. The fed-back voltage will then, as we have seen, tend to damp out the signal.

If, on the other hand, the applied frequency (or alternatively the capacitance of C) is reduced, more current will flow through L than through C, giving us an *inductive* anode circuit. Now the coupling between the two tuned circuits provided by C_{ga} will feed back energy that assists and builds up the voltage already present.

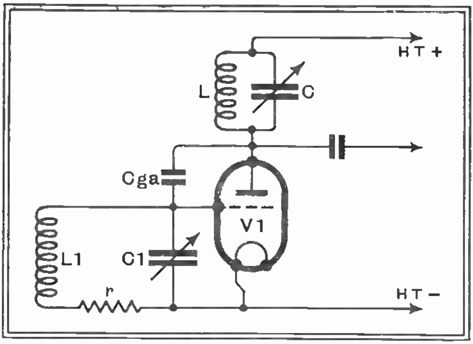


Fig. 102: The grid-anode capacitance of V_1 introduces difficulties into the working of the tuned-anode circuit

In discussing reaction (Sec. 101) we saw that energy fed back into a tuned circuit could be made to reduce the effective resistance of that circuit almost to zero by supplying energy almost as fast as it was dissipated in the natural circuit resistance r. Suppose we feed back energy *faster* than it is being used up, making the effective resistance of the grid circuit *negative*.

108. Instability

If this happens, any slight current present in L_1C_1 will grow by virtue of this excess energy, and will go on growing as long as the valve continues to feed back more energy than is dissipated in r. Since rising volts on the grid produce proportionately rising volts on the anode, the current in L_1C_1 will continue to increase until this proportionality breaks down, which will only occur when the voltages are so large as to enter upon the non-linear part of the valve's characteristics. Then the average slope of the valve will be reduced and grid current increased until the energy fed back is only just sufficient to replace that lost in r, and a state of equilibrium will be attained.

The valve is now said to be *oscillating*. It is producing and maintaining in L_1C_1 a constant alternating current at the frequency to which this circuit is tuned, this current producing across L_1C_1 a voltage at least equal to the largest that the valve can handle without distortion.

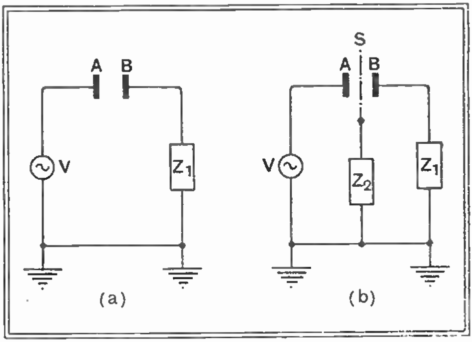


Fig. 103 : Illustrating the theory of screening

If C_{ga} is large enough, if r is small enough, and if this amplification afforded by the valve is great enough, this is what happens in the circuit of Fig. 101. With coils of fairly good design (low r) and any ordinary triode, oscillation appears every time an attempt is made to bring L_1C_1 and LC into resonance with the same frequency. Although theoretically there should be no tendency to oscillation when exactly tuned, it is found that the increasing loudness of signals due to the commencement of feedback as C is reduced below the value necessary for resonance completely overwhelms the decrease of loudness that one would expect to find on detuning. In tuning the set there is therefore no aural indication of the true resonance point, so that in trying to tune for loudest signals one is led, every time, straight into the trap of oscillation, which occurs as soon as C is set a fraction low in capacitance.

In a receiver, oscillation results in the production of a rushing noise, and in the development of sundry whistles and squeaks as the set is tuned. These are not merely supplementary to the musical programme required ; they replace it. For all practical purposes, therefore, the circuit of Fig. 101 is unusable.

When the triode was the only valve available, oscillation due to feed-back through C_{ga} was avoided by providing a "faked" circuit by means of which another voltage, equal in magnitude but opposite in phase to that causing oscillation, could be fed back to the grid of the valve. These arrangements were known as *neutralized* circuits. They have now died out entirely, the modern solution to the problem of preventing feed-back through the grid-anode capacitance of the valve lying in the choice of a valve in which, by internal screening, this capacitance has been reduced practically to zero.

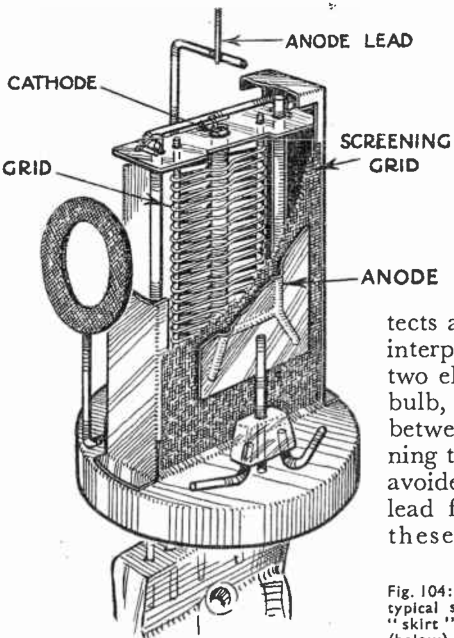
109. The Theory of Screening

The capacitance between any two objects can be reduced to zero by interposing between them as a screen an earthed metal sheet of sufficient size. The operation of such a screen can be understood by considering Fig. 103 which shows at *a* two plates A and B separated from one another by an air-space. There will be a capacitance between them, so that the radio-frequency generator V will drive a current round the circuit Earth—V—A—B—Z_1—Earth. Across Z_1, which is an impedance of some kind between B and earth, the current will develop a potential difference, and this P.D. will be the voltage appearing on B as a result of the passage of current through the capacitance AB.

At *b* a third plate S, larger than either of the two original plates, is inserted between them in such a way that no part of either plate can "see" any part of the other. We now have no direct capacitance between A and B, but we have instead two capacitances, AS and SB, in series. If an impedance Z_2 is connected between S and earth the current round the circuit Earth—V—A—S—Z_2—Earth will develop a P.D. across Z_2. Since Z_2 is also included in the right-hand circuit the P.D. across it will drive a current round the circuit Earth—Z_2—S—B—Z_1—Earth, and this will give rise to a potential on B. So far, S has not screened A from B, there remaining an effective capacitance between them which, if Z_2 is infinitely large, amounts to the capacitance equivalent to that of AS and SB in series. If S is thin this is practically equal to the original direct capacitance between the two plates.

Now imagine Z_2 to be short-circuited. Current will flow round the first circuit, but since there is now no impedance common to both there will be no driving voltage to produce a current in the latter. No matter what alternating voltages are applied to A, none will appear on B, even though large currents may flow via S to earth. The effective capacitance between A and B has therefore been reduced to zero, and B is completely screened from A.

It is very important to note that S is only effective as a screen if it entirely cuts off A from B, thus replacing the direct capacitance AB by AS and SB in series. Even with this proviso, perfect screening is not obtained unless S is definitely connected to earth either by a direct wire or through an impedance Z_2 which is negligibly small.



110. Screening a Valve

This is the principle used in reducing the grid-anode capacitance of a valve. A screen, so designed that it completely protects anode from grid, is interposed between these two electrodes within the bulb, while capacitance between the leads running to grid and anode is avoided by taking the lead for one or other of these electrodes out

Fig. 104: Showing construction of a typical screened valve. Note the "skirt" screening the grid lead (below) from the anode. This "skirt" is connected to the screen

through the top of the bulb. In some recent types of valve the screening has been extended down to and below the base, allowing both anode and grid to be brought out at the same end.

Clearly, a solid metal screen, while providing irreproachable screening, would cut off the electron flow from cathode to anode ; it is therefore necessary to use as screen a close-mesh wire gauze through the openings of which electrons can pass. It is found that this necessary compromise with perfection still leaves a completeness of screening that falls short of that obtainable with an unbroken sheet of metal by a surprisingly small amount. In an unscreened valve, C_{ga} is usually of the order of 6 to 8 $\mu\mu$F. ; with a gauze screen, properly earthed, this is commonly reduced to less than 0·003 $\mu\mu$F, and may even be less than 0·001 $\mu\mu$F. The structure of a typical screened valve is shown in the sketch of Fig. 104.

111. How a Screened Valve Works

If earthed in the strictly literal sense the potential of the screen would be approximately that of the cathode. Since the attraction of the positive anode cannot extend through the screen to any appreciable extent, electrons in the neighbourhood of the grid of the valve would then not be drawn onwards, and the anode current would fall practically to zero. But since, as Fig. 103 shows, the requirements of screening can be met by making Z_2 negligibly small, we can connect a condenser of large capacitance from the screen of the valve to earth, after which we can supply the screen, from any convenient source, with a positive potential.

The inner portion of the valve, comprising cathode, grid, and screen, is practically unaffected by the voltage at the anode ; in consequence the total current through the valve is almost completely determined by the potentials of grid and screen. But if an electron arriving at the screen should happen to find itself exactly opposite to one of the openings in the latter, the attraction exerted upon it by the screen will come equally from all sides and it will go straight through the opening. With the anode at zero potential it would fall back again to the screen, but if the anode is much more positive than the screen it will be drawn on.

Thus by making the anode more positive than the screen some of the electrons, initially set in motion by the positive potential on the screen, will pass through the latter and travel on to the anode. The more the potential of the anode exceeds that of the screen the more electrons will be drawn on; with rising anode voltage, therefore, the anode current rises and the screen current falls, the total remaining practically constant.

112. Characteristics of a Screened Valve

Curves of a typical screened tetrode are reproduced in Fig. 105, which shows anode current plotted against anode voltage. Each curve refers to the fixed grid-voltage E_g mentioned against it, and all were taken at a fixed screen-voltage of $E_s = 80$ v. So long as E_a is considerably more than E_s, the anode takes practically all the current; over the range $E_a = 120$ to $E_a = 200$ v. on the curve for $E_g = -2$, the anode current changes by only 0·08 mA. As E_a falls below 120 v. the proportion of electrons pulled through the screen to the anode begins to drop, as the rapid fall in I_a shows. The screen current I_s, if plotted, would show a corresponding rise, keeping the total space-current constant.

The reasons for the peculiar shape of the curves for values of E_a lower than E_s will be discussed in connection with pentodes; for the present it is enough to note that a screen-grid valve is always used with an anode voltage considerably higher than that on the screen.

The extreme flatness of the curves over the working region to the right of the diagram indicates that the anode resistance of the valve is very high (Sec. 64). For the curve $E_g = -2$, the change of I_a by 0·08 mA for a change in E_a of 80 v. indicates a resistance of $80/0{\cdot}00008 = 1$ megohm. But this value depends far more than in the case of the triode upon operating voltages. Reducing the grid bias reduces also the anode resistance; reading off values from the curve for $E_g = -1$ gives an anode resistance of

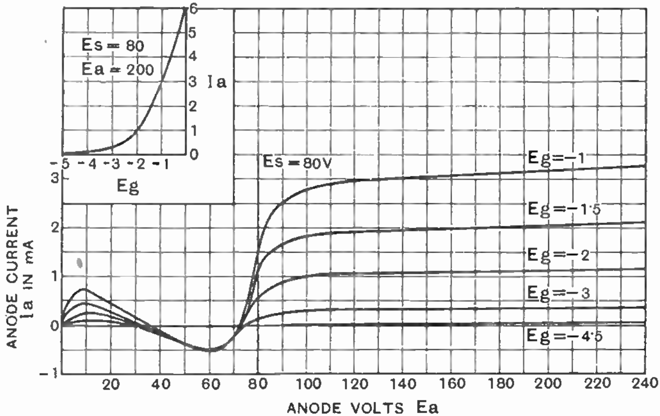


Fig. 105: Characteristic curves of typical screened tetrode. Only the flat parts of the curves to the right of the line E_s are used for amplification. Inset : I_a—E_g curve to show slope

350,000 ohms only, which is about one-third of the value found for $E_g = -2$.

The small curve inset on Fig. 105, which shows the variation of anode current with grid voltage at $E_s = 80$ and $E_a = 200$, makes clear that this rapid variation of anode resistance is not accompanied by corresponding changes in mutual conductance or slope. At $E_g = -1$, $g_m = 2{\cdot}45$, while at $E_g = -2$, $g_m = 1{\cdot}45$ mA/v. Since the amplification factor of the valve is given by $\mu = g_m r_a$, we can find its value from the figures for g_m and r_a at these two bias points ; at $E_g = -2$, $\mu = (1{\cdot}45/1000) \times 1{,}000{,}000 = 1450$, while at $E_g = -1$, $\mu = (2{\cdot}45/1000) \times 350{,}000 = 880$.

In the triode, the amplification factor is determined almost entirely by the geometry of the valve, and therefore does not vary over these considerable ranges ; further, it is much lower, seldom exceeding 100. Nevertheless, the screen-grid valve, used as a radio-frequency amplifier, does not give such enormously enhanced gain as these startingly high figures might suggest, for their effect is very largely offset by the valve's very high anode resistance.

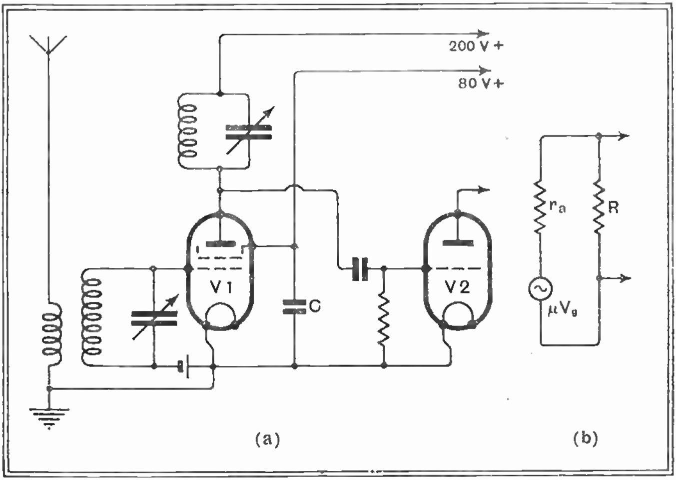


Fig. 106 *a*: A simple R.F. stage employing a screened valve with tuned-anode coupling; *b* equivalent anode circuit of the valve. If R is small compared with r_a, gain of stage is approximately $g_m R$

113. Finding the Gain

Fig. 106 *a* shows a simple tuned-anode stage of radio-frequency amplification, preceding a grid-detecting triode V2; with the exception of the addition of the screen circuit, with its large by-pass condenser to earth, the arrangement is the same as that for a triode. At *b* is shown the equivalent anode circuit of the valve, the signal-voltage V_g at the grid being represented, as before, by μV_g volts in series with the anode resistance of the valve. If R, the dynamic resistance of the tuned circuit, is 100,000 Ω, the amplification given by the stage, calculated from the formula $A = \frac{\mu R}{R + r_a}$, works out as 196 times for $E_g = -1$ and 141 times for $E_g = -2$. The rising amplification factor has been accompanied by so large a rise in anode resistance that the gain actually *drops* in passing from $E_g = -1$ to $E_g = -2$.

In most practical cases the resistance of the valve is so very much higher than that of the tuned circuit connected to its anode that R is small compared with r_a. A good approximation to the correct value for the stage-gain can then be had by writing $A = \mu R/r_a$, or $A = g_m R$.* The conditions for high gain with a screen grid valve are therefore simply that we choose a valve of high slope and follow it with a tuned circuit of high dynamic resistance.

Apart from these considerations the screen-grid valve behaves exactly like a triode from which the grid-anode capacitance has been removed ; all the principles and methods discussed in Chapter 7 can therefore be applied to the tetrode.

114. The Limits of Stable Amplification

The introduction of the screening makes it quite possible to build up and use successfully a circuit such as that of Fig. 106 *a* without running into difficulties due to oscillation. It can be shown that the stage will be stable provided that the numerical value of a quantity H is less than 2. This quantity is given by the relation $H = 2\pi f g_m C_{ag} R_1 R_2$, where f is the frequency of the signal being amplified, and R_1 and R_2 are the effective dynamic resistances of the tuned circuits connected to grid and anode. High values of R_1 and R_2, which imply circuits of low inherent losses, tend, as might be expected, to produce oscillation. So also do high values of valve-slope or grid-anode capacitance, while the likelihood of instability is greater, other things being equal, the higher the frequency of the signal it is desired to amplify.

For a valve for which $g_m = 2{\cdot}5$ mA/v, $C_{ag} = 0{\cdot}005\ \mu\mu F$, used at 1,500 kc/s (200 metres), we can find now the maximum dynamic resistance that the tuned circuits can have without causing oscillation. For critical oscillation $H = 2$, so that we can write $R_1 R_2 = \dfrac{2}{2\pi f g_m C_{ag}} = \dfrac{2}{118} \times 10^{12}$.
If the two tuned circuits are alike each may have a maximum dynamic resistance equal to the square root of this ; i.e., of 130,000 ohms. Since this represents

* g_m in amps per volt and R in ohms, *or* g_m in milliamps per volt and R in thousands of ohms.

a tuned circuit only a little better than the average, it is clear that the inter-electrode capacitance assumed for the valve is just on the maximum permissible limit for a single stage of amplification. In such a case the amplifier, though just stable, will be quite near oscillation, and we have a condition in which feed-back through the valve is not far from sufficient to reduce the radio-frequency resistance of the grid circuit to vanishing point.

115. Two Stages

For reception of the most distant stations, the gain given by a single stage of amplification is hardly adequate, and it is desirable to add a second. This brings up, in much more acute form, the difficulty of instability and experience shows that it is very difficult to persuade two tuned-anode stages to refrain from self-oscillation.

Examination of the two-stage tuned-anode amplifier of Fig. 107 shows that the tuned circuit 2, besides being in the anode circuit of V_1, serves as grid circuit for V_2, being connected between the grid of that valve and the H.T. line. This, being at zero potential so far as signals are concerned, counts as "earth" from the A.C. point of view. To keep the grid of V_2 at the right potential for amplification, grid bias is connected through a resistance high enough not to damp circuit 2 excessively. In its capacity of grid-circuit to V_2, the tuned circuit has energy fed into it through the valve, and so has its R.F. resistance reduced well below its normal value. This results in giving it a very high dynamic resistance, and it is this artificially-raised figure that must be taken for R_2 in applying the formula to compute the stability of the first stage. As the formula shows, a rise in R_2 increases the tendency to oscillation, and we conclude that two stages, each individually stable, may oscillate if connected in cascade as in Fig. 107.

116. Transformer Coupling

Feedback from the anode of V_1 to its grid can be reduced by cutting down the signal-voltage at the anode. Naturally one dislikes sacrificing gain, so that one would like to maintain as nearly as possible the signal-voltage eventually reaching the grid of V_2. This can best be done by

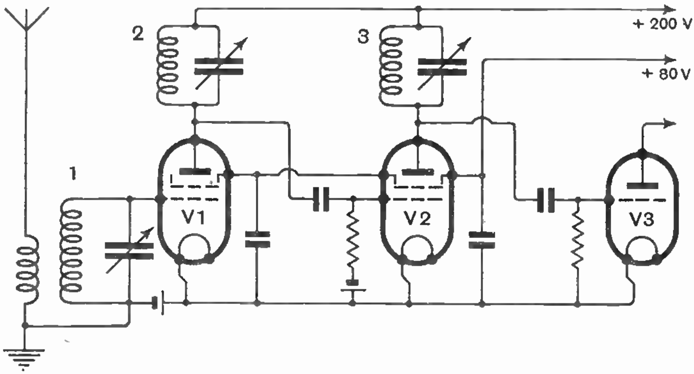


Fig. 107: Two-stage radio-frequency amplifier, using tuned-anode circuits with screen-grid valves

replacing each tuned circuit with a radio-frequency transformer of the conventional type, using a tuned secondary and an untuned primary. The conversion of Fig. 107 to this more stable arrangement is shown completed in Fig. 108. The exact turns-ratio that gives best results in such a case is usually best found by experiment, but the gain can readily be computed for any ratio to which a search for stability may lead us.

If the secondary has a dynamic resistance R, and contains n times as many turns as the primary, the effective resistance of the latter will be R/n^2. Following a valve of slope g_m, the signal-voltage at the anode will therefore be g_mR/n^2 times that at the grid, while at the grid of the succeeding valve it will be n times this owing to the voltage step-up in the transformer. This makes the gain, reckoned from grid to grid, equal to g_mR/n.

Thus if we replace a tuned-anode coupling, the gain for which is g_mR, by an R.F. transformer of ratio n, we divide the stage-gain by n and the voltage at the anode of the valve by n^2. Thus we can cut down the signal-voltage at the anode to one-ninth of its value in the simple tuned-anode circuit at the cost of dividing the gain of the stage by only three.

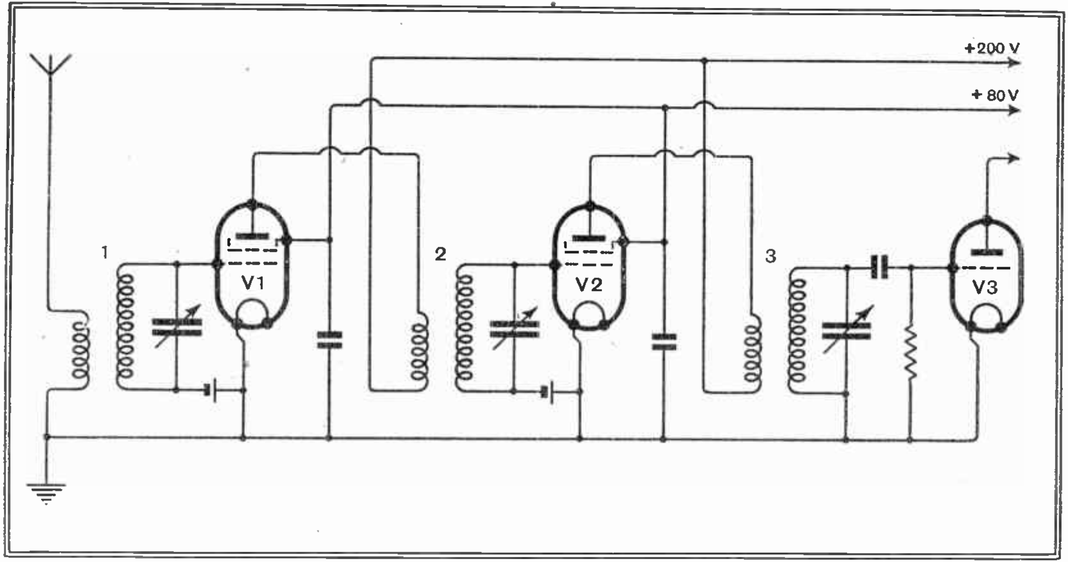


Fig. 108 : Two-stage radio-frequency amplifier, using step-up transformer couplings with screen grid valves. Much more stable than the closely-corresponding circuit of Fig. 107

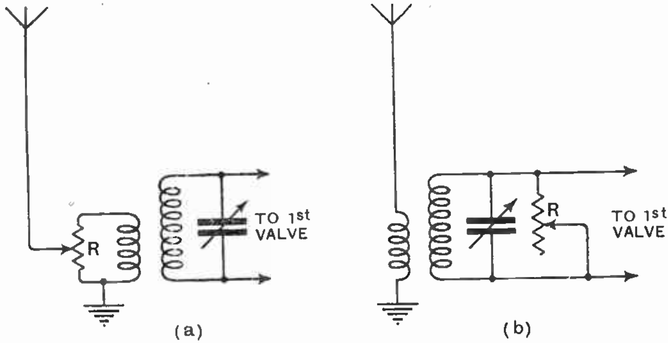
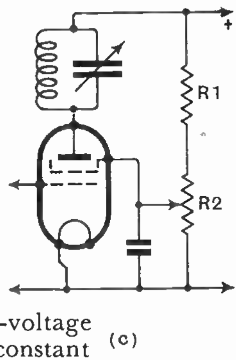


Fig. 109 : Three general methods of controlling the signal passed to V_3 in Fig. 108



117. Volume Control

In order to prevent the detector-valve V_3 from being grossly overloaded when receiving a near-by station, it will be necessary to add to the circuit of Fig. 108 some form of *volume control* by manipulation of which the overall gain of the amplifier can be adjusted. By this means it is possible to ensure that the signal-voltage reaching the detector is kept at a constant value irrespective of the voltage produced at the aerial by the particular transmitter tuned in.

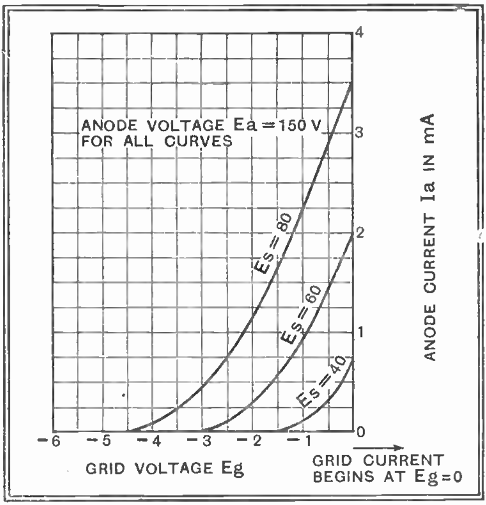
Volume control can be obtained in the three general ways illustrated in Fig. 109; by controlling the input from the aerial, as at *a*, by controlling the magnification of one or more tuned circuits, as at *b*, or by controlling the gain given by the valve, as at *c*. With method *a* the amplifier works always at full gain, in which condition it is likely to produce a certain amount of background noise ("valve-hiss") which, while tolerable in listening to a distant station, must be avoided, if possible, while listening to a near one.

For this reason method *a* is not used save as an auxiliary to some other type of control.

Method *b* suffers from the drawback that in reducing the gain of a tuned circuit its selectivity is reduced also ; save for local-station reception, where this is sometimes considered an advantage, this type of control is not used.

Method *c* is, theoretically, ideal, since it supplies a means of controlling gain by reducing the amplification given by the valve, the slope of which drops as E_s is decreased, without affecting any of the other characteristics of the amplifier. In the particular form shown in the diagram, however, it leaves a good deal to be desired, as can be seen by reference to the curves of Fig. 110.

Fig. 110 : Curves of ordinary screen-grid valve. Note that rectification (overload) can occur on quite a small signal, especially when E_s is reduced



118. Distortion due to the R.F. Amplifier

Here are shown the E_g — I_a curves of a typical screen-grid valve, and it is at once evident that when the voltage on the screen is lowered the available portion of the characteristic, lying between the grid-current region and cut-off, is neither long enough nor straight enough to accommodate a signal of any but very small magnitude. As always, a curved characteristic means rectification, with its accompanying distortion, and it is clear that with such a volume control as this, distortion will be greatest where we can least tolerate it—when receiving the local station.

If the valve were dealing with a simple unmodulated carrier distortion would be harmless, for distortion of a simple waveform means no more than that there are added to it various harmonics. Since subsequent tuned circuits, tuned to the fundamental frequency, would not respond to these, they could never reach the detector-valve, and so no harm would be done.

Unfortunately, our valve has to deal with a modulated wave ; in other words, with a whole spectrum of closely-related frequencies. Distortion of such a complex signal results in the importation into the signal of new sidebands which are removed from the carrier two and three times as far in frequency as the original sidebands from which the valve produced them. Beyond the detector, these appear as harmonics of the note originally transmitted.

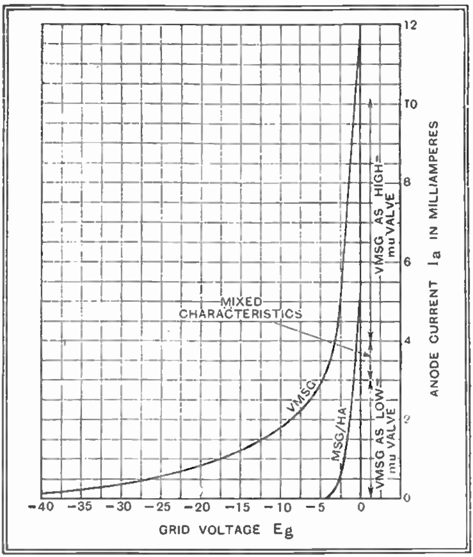
119. Cross-Modulation

Besides this distortion of a single modulated carrier there is a type of distortion, known as *cross-modulation*, which makes its appearance under the misleading guise of lack of selectivity. It arises like this. Suppose that the receiver of Fig. 108 is tuned to a station 45 kc/s away from the local. We may very well assume that the overall selectivity of the three tuned circuits is enough to reduce the local station to inaudibility when they are all tuned 45 kc/s away from it. But the grid of the first valve is only protected from the local station by one single tuned circuit ; it is not impossible that at this grid this station may produce quite a large voltage. If this voltage is large enough to cause the valve to rectify, one family of the resultant valve-produced frequencies consists of the carrier of the station to which the set is tuned modulated with the programme of the local station. Since the set is tuned to this carrier, the remaining two tuned circuits will pass it along, together with its twin programmes, to he detector, which will make both stations audible ogether. If the station to which the set is tuned switches ff its carrier wave the programme of the local station will lso disappear, thereby proving beyond all doubt that the nterference is due to cross-modulation, and not simply o lack of selectivity in the tuned circuits.

For all practical purposes, the selectivity of a set in vhich cross-modulation is occurring is no greater than hat of the tuned circuit preceding the first grid. In sets of this type it is therefore common practice to interpose wo tuned circuits between the aerial and the first valve.

For more satisfactory prevention of cross-modulation ve shall have to replace the first valve with one which overloads less readily, so that quite large voltages from he local station can reach it without causing rectification. Further, this new valve must be suited to some means of gain-control other than that obtainable by eduction of creen voltage, which must inevitably reduce he signal-acceptance of he valve.

ig. 111 : Characteristic of variable-mu valve (VMSG) ompared with that of standard screen-grid valve. Note he increased signal-handling ability of the VMSG and the slow but steady change of slope with bias



120. The Variable-Mu Tetrode

To fulfil hese conditions he variable-nu screen grid valve has been produced. It differs from the

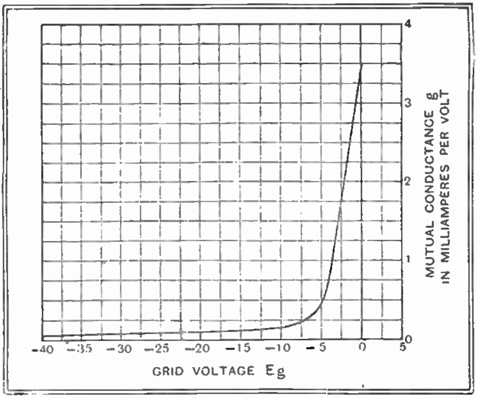


Fig. 112: How the mutual conductance of a variable-mu valve (VMSG) is affected by alterations of bias

normal screen-grid valve only in having a control-grid with a mesh of uneven pitch. Where the mesh is close it behaves as an ordinary screened valve of high amplification factor; where it is open, the flying electrons are controlled as in a low-mu valve, and a large negative bias is consequently required to reduce the anode current to zero. The valve behaves as, and in effect actually is, two valves in parallel.

In Fig. 111 is plotted the $E_g - I_a$ curve of a variable-mu valve, the curve of an ordinary screen-grid valve being plotted, for comparison, on the same diagram. As the curve at once shows, the high-mu component of the variable-mu valve is effective at low bias values, while at high bias the low-mu portion alone is in operation, since the electrons are unable to penetrate the close-mesh portion of the grid when this is very negative.

The value of this valve does not only lie in the fact that it has a characteristic long enough to accommodate a very strong signal without serious distortion; in addition, the change of mutual conductance with bias allows us to use bias variations as a means of controlling amplification. We have already seen (Sec. 113) that the gain given by a screened valve is approximately proportional to the slope; Fig. 112 shows how this varies with applied bias, and makes clear how, by increasing the bias, the gain of the stage can be reduced to almost any desired extent.

The curve (Fig. 111) is still not straight, so that distortion and cross-modulation are still theoretically possible. In practice, their appearance when using a variable-mu valve is a rarity, because it is extremely seldom that a received signal is strong enough to sweep the grid over more than a very small portion of its characteristic—and over any very small range this curve, or any other, may be regarded as substantially straight. And it must be remembered that for strong signals the bias is increased—primarily for the sake of reducing gain, but incidentally providing a working point suited to a strong signal.

As a result of these advantages over the simple screen-grid valve, the latter has been almost entirely ousted. In no other respect than those just touched upon is there any difference between the two types of valve ; with the obvious minor modifications, all that has been said about the simpler valve may be applied, without alteration, to its successor.

121. Secondary Emission

While the introduction of variable-mu characteristics overcomes with fair completeness overloading and distortion arising in the grid circuit of the valve, there remain possibilities of trouble in the anode circuit. These arise owing to the peculiar shape of the $E_a - I_a$ curve, which is shown in full line in Fig. 113. If the sole effect of raising the anode voltage were to rob the screen of more and more electrons, the valve curves would take a form such as that shown dotted on the same diagram. Why the divergence between theory and observed fact ?

As always when theory and practice do not agree, the theory has overlooked something. In the present case it has omitted to take into account the phenomenon of *secondary emission*, by which is meant the ability of a fast-moving electron to knock out another electron when it strikes a metal surface. Once liberated, free electrons so produced will naturally be attracted to the most positively charged object in their neighbourhood.

At low anode voltages the real curve follows the dotted one, but at A the velocity of the electrons has risen enough to enable them to dislodge secondary electrons from the anode on their arrival there. These electrons find their

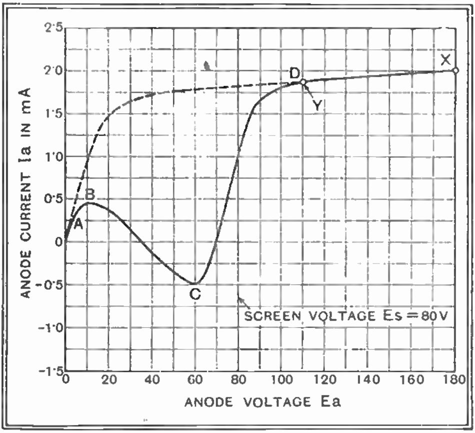


Fig. 113: "Theoretical" (dotted) and actual (full line) curves of tetrode valve. The extraordinary shape of the latter between A and D is due to secondary emission from the anode. The introduction of a "suppressor" grid, turning the valve into a pentode, enables the dotted curve to be realized in an actual valve

way to the more positive screen, so reducing the net number of electrons arriving at the anode, and reducing the anode current below the "theoretical" value. Beyond B, the peak of the curve, each extra electron drawn to the anode by rising voltage knocks out more than one when it gets there, and these all reach the screen, which still has the higher potential. The current, therefore, *decreases* with rising anode voltage. It even reverses in direction, this merely meaning that the total number of electrons arriving at the anode is less than the number they dislodge by secondary emission.

At higher anode voltages than that at C, the secondary electrons begin, in increasing numbers, to return to the anode, allowing the anode current, therefore, to begin to return towards its "theoretical" value. Finally, as soon as E_a exceeds E_s by a small amount (at D) the superior attraction of the anode prevents any from reaching the screen. The observed curve has now joined the dotted curve, showing that secondary emission no longer has any effect on the net anode current.

Secondary emission, although it must occur, does not distort the characteristic curves of a triode valve, for the excellent reason that secondary electrons, when emitted, always return to the anode, since it is the only positively harged object near them. The total anode current is ıus not altered by their temporary absence from the node.

Consideration of the full-line curves in Fig. 113 makes : perfectly clear that if the voltage at the anode is swung y the signal so far that it falls momentarily down to ıat of the screen, violent distortion is likely to occur. f, for example, $E_a = 180$ v. and $E_s = 80$ v., the maximum ermissible signal swing at the anode is about 70 volts eak (from X down to Y) ; after that, rapid curvature egins.

122. The Screened Pentode

Admittedly, signal voltages of this order are seldom equired in a radio-frequency stage, so that distortion of his type does not often occur. Nevertheless, its source an be removed by inserting between screen and anode n extra grid, connected to cathode, which will serve to rotect the electrons dislodged from the anode from the ttraction of the screen, so ensuring that, as in the case f the triode, they all return to the anode. This extra rid is called a *suppressor grid* by virtue of the fact that it suppresses" secondary emission, and a valve containing t, having five electrodes, is known as a *pentode*. The shape f the $E_a - I_a$ curves of the pentode is, as theory predicts, ractically that of the dotted curve of Fig. 113.

Like the screened tetrode, the screened pentode is available in both variable-mu and short grid base types ; the ormer is intended primarily for R.F. amplification, while he latter makes a serviceable detector or audio-frequency mplifier. The addition of the suppressor still further educes the influence of the anode in determining the otal space-current through the valve ; in other words, he pentode has a higher anode resistance (and consequently a higher amplification factor) than a corresponding etrode of the same slope. Since, in a radio-frequency mplifier, the valve is shunted across the tuned circuit (as in ig. 107), this high anode resistance results in a slight gain n selectivity as compared with the tetrode ; save for this ne point, and the total elimination of the possibility of anode-circuit overload through running the anode to voltage less than that of the screen, the screened pentod and the screened tetrode may be regarded as identical Except when overloading is possible, no difference whatever will be found, in practical use in a receiver, betwee the two valves.

It may be useful to give here a summary of the outstanding characteristics of each of the types of valves s far discussed.

Diode : Two electrodes (cathode and anode) only. Rectifies, but will not amplify.

Triode : Cathode, grid, and anode. Amplifies, oscillates and detects. Is the fundamental type of valve, fron which more elaborate structures have developed.

Screened Tetrode : A triode with addition of a screen betwee anode and grid to prevent instability. Anode resistanc and amplification factor very high.

Variable-mu Screened Tetrode : As preceding, but with grid-circuit overload reduced and adequate means of gain control provided.

Screened Pentode : As screened tetrode, but capable o dealing with large signal at anode.

Variable-mu Screened Pentode : Combines advantages of both the two preceding valves. The most developed type which has now almost completely ousted the precedin three.

CHAPTER 12

SELECTIVITY IN THE R.F. AMPLIFIER

123. Resonance Curves

IN trying to raise the sensitivity of a simple single-valve set we first tried reaction, using it to reduce enormously the radio-frequency resistance of our simple set's one tuned circuit. The terribly over-sharp tuning and consequent loss of sidebands that accompanied this attempt led us to reject it in favour of obtaining amplification by the aid of additional valves as radio-frequency amplifiers. We then found that to make these amplify satisfactorily we had to introduce extra tuned circuits. The question at once arises whether, in adding these extra circuits, we have not committed ourselves to just as great an accentuation of selectivity as we originally got with a single circuit and reaction. To settle this point we shall have to go a little more deeply into the subject of resonance curves, both of single circuits and of several in combination.

From the point of view of the adequate reception of high notes, all we need to know is the amount by which the response of our tuned circuit drops at a frequency removed from resonance by the frequency of the musical note we wish to consider. This depends solely on the ratio of the inductance of the coil to the radio-frequency resistance of the tuned circuit as a whole. For all wireless problems, we are only concerned with the response at frequencies not very far removed from resonance, for finding which the formula that follows, although a little simplified, is amply accurate.

If the voltage across the tuned circuit at resonance is

V_o, and that across it for a frequency n cycles from resonance is Vn, then $V_o = Vn\sqrt{1 + (4\pi n)^2\left(\frac{L}{r}\right)^2}$.

The complete square root, which we will hereafter abbreviate to s (for selectivity), tells us by how much we must multiply the voltage at n cycles from resonance to get the voltage at the resonant point. If $s = 4$ at 10 kc/s off tune, the voltage at this frequency is one-quarter o that at resonance, and we speak of the circuit as being "four times down at 10 kc/s off tune."

The expression for s is rather a troublesome one to evaluate quickly for a rapid comparison of the selectivity of different circuits; actual values of s are therefore shown for L/r ratios up to 500 in the curves of Fig. 114. Separate curves are given for 5, 9, 18 and 27 kc/s off tune.

124. Reaction and Amplification Compared

With the aid of these curves we are in a position to compare at once the selectivity of a reacting detector with that of a set containing a single stage of radio-frequency amplification and therefore employing two tuned circuits. If we assume that at some particular frequency the ratio L/r of the tuned circuits in the amplifier is 10, then we see from Fig. 114 that at 5 kc/s off tune each circuit has its response reduced to $1/1{\cdot}18$ of that at resonance. For two tuned circuits the overall response will be the square of this, or $1/1{\cdot}39$; that is, the amplifier will pass 72 per cent. of the side-bands representing high notes of frequency 5,000 cycles.

If the gain of the stage is assumed to be fifty times, then to get equal amplification by means of reaction we shall have to reduce r to one-fiftieth of its normal value, thereby increasing L/r to 500. Reference to Fig. 114 shows that with L/r raised to this value s becomes 30, making the response at 5 kc/s off tune one-thirtieth that at resonance. In this one tuned circuit side-bands are so cut that only some 3 per cent. of a 5,000-cycle note will reach the speaker. The loss at this frequency is thus some 24 times as great as when using the extra tuned circuit necessitated by the valve, although the amplification afforded is in each case the same.

125. Separating Stations

Selectivity is often regarded as the ability of the set—which means of the tuned circuits in it—to select one station to the exclusion of others. In allotting wavelengths to the various transmitting stations, international agreement has resulted in a uniform spacing, from each station to the next, of 9 kilocycles per second. The 9 kc/s gap between carriers is left to cover the " spread " of frequency taking place as a result of modulation of the carriers by the programme.

If one station transmits at 191 kc/s, its two neighbours will transmit at 200 and 182 kc/s respectively. The wavelengths corresponding to these frequencies are, in order, 1500, 1571 and 1648 metres, making an average spacing between stations of 74 metres. If we consider stations transmitting at much higher frequencies, the same 9-kc. separation holds, because the width of sidebands is determined by the audible *frequencies* in the programme, and has nothing to do with the wavelength of the carrier. Three stations in order from the list transmit on 1474, 1465 and 1456 kc/s : expressed in wavelengths, these frequencies are equivalent to 203½, 204¾ and 206 metres, a spacing between stations of 1¼ metres.

These figures make it abundantly clear that separation between stations cannot intelligibly be expressed in metres ; a proud boast that " My set will separate stations only 20 metres apart " means nothing at all unless there is also specified the wavelength at which this prodigy of selectivity (or woeful lack of it, as the case may be) was observed. We shall therefore have to deal with selectivity exclusively in terms of frequency. The figures further show that we shall not wish to be concerned with the actual carrier-frequency in use ; all that concerns us is the amount by which the reponse of our tuned circuit drops at some known number of kc/s from resonance.

We shall find it convenient to use, therefore, the formula and curves already discussed in considering quality.

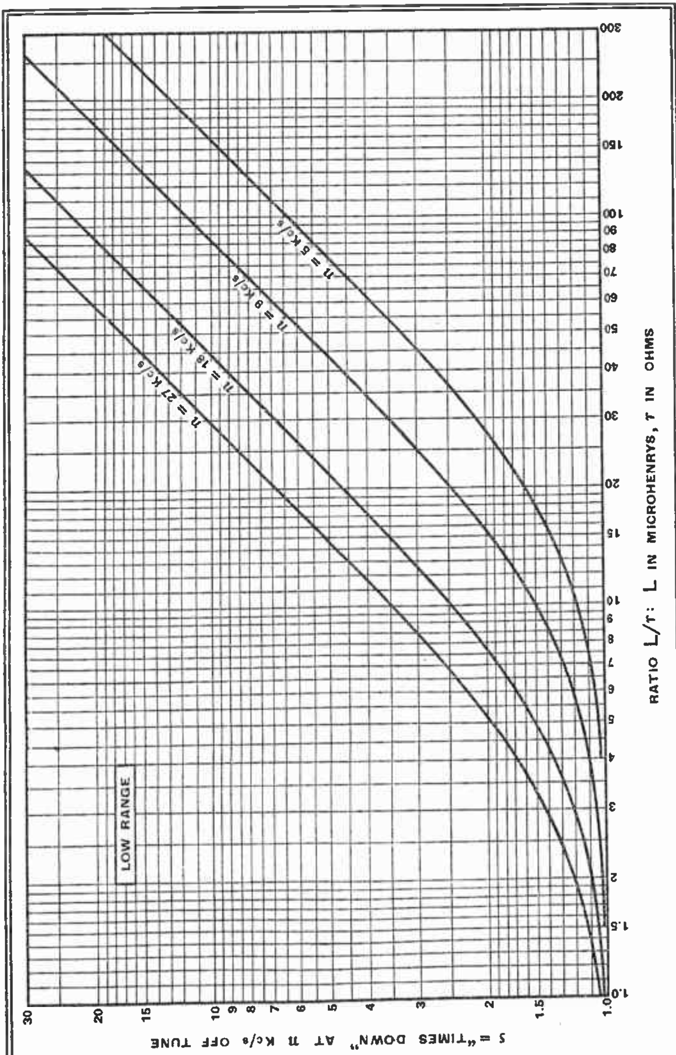


Fig. 114a: Showing relationship for one tuned circuit between L r and s at 5, 9, 18, and 27 kc s off tune. Where value of s is greater than 10, use the continuation of these curves in Fig. 114 b

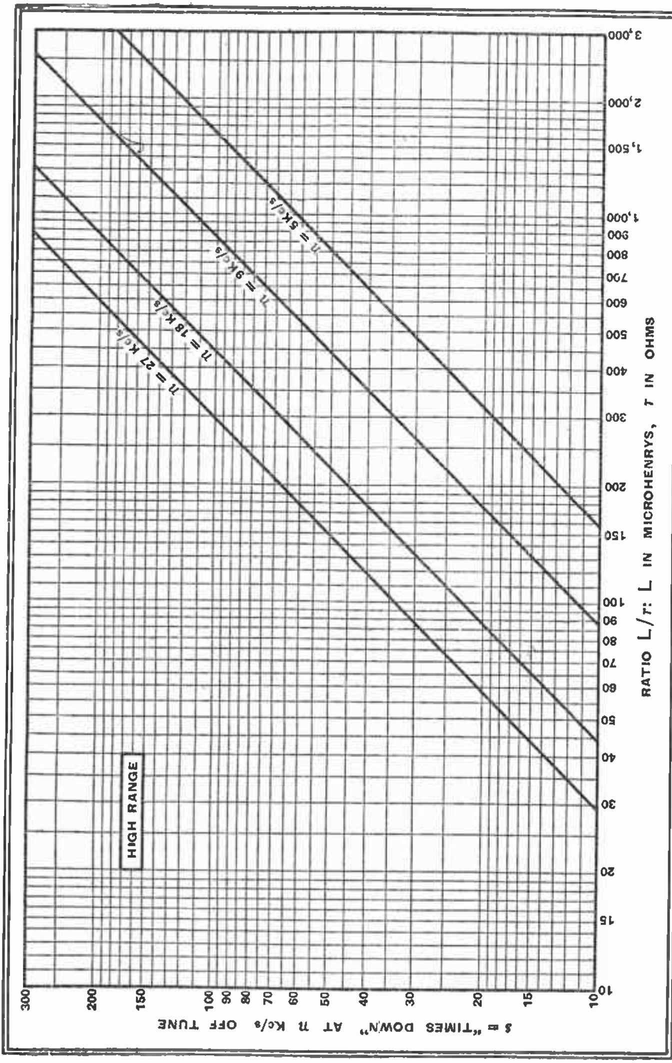


Fig. 114 b : Continuation, to higher values, of the curves of Fig. 114 a

Since we would like to retain at full strength frequencies off tune by at least 5 kc/s for the sake of quality, and yet, for the sake of selectivity, would like to remove as completely as possible frequencies 9 kc/s off tune, we require, if we can get it, a resonance-curve with a flat top and steeply-falling sides.

126. Conflicting Claims

Some approximation to this can be obtained by using a large number of fairly flatly tuned circuits in cascade. Where a number of circuits are so used the overall s is found by raising the s-value for one circuit to the appropriate power—squaring for two circuits, cubing for three, and so on. To enable the reader to find for himself the behaviour of any series of tuned circuits in which he may be interested, Fig. 115 gives curves in a rather more general form than Fig. 114. In place of plotting s against L/r, and making a separate curve for each value of n, s is here plotted against the product $n \times L / r$. Curve 1 refers to one tuned circuit, curve 2 to two circuits, and so on up to a total of six circuits, all connected in cascade.

To find, for example, " times down at 9 kc/s " for a series of circuits for each of which $L / r = 10$ we only have to multiply 10 by 9 to find $n \times L / r$, and look up the required figure on the curve corresponding to the number of tuned circuits for which the result is required. For one tuned circuit we find that $s = 1{\cdot}5$, for two $2{\cdot}25$, for three $3{\cdot}38$, and so on. Alternatively, to find the requisite L / r to give 10 times down at 9 kc/s with four circuits, the value of $n \times L / r$ corresponding to $s = 10$ is read off from the curve for four circuits, and is found to be 117. The required L / r is then $117 / 9 = 13.0$.

127. Equal Selectivity

Suppose, for example, we require to reduce the voltage of an interfering station 18 kc/s off tune to one-hundredth of the voltage it would have if exactly tuned in. As Fig. 114 shows, a single circuit to do this has $L/r = 450$, with which a 5 kc/s side-band will be 28 times down. If we used six tuned circuits the value of $n \times L / r$ required, as Fig. 115 shows, is 152, giving $L / r = 152 / 18 = 8{\cdot}4$.

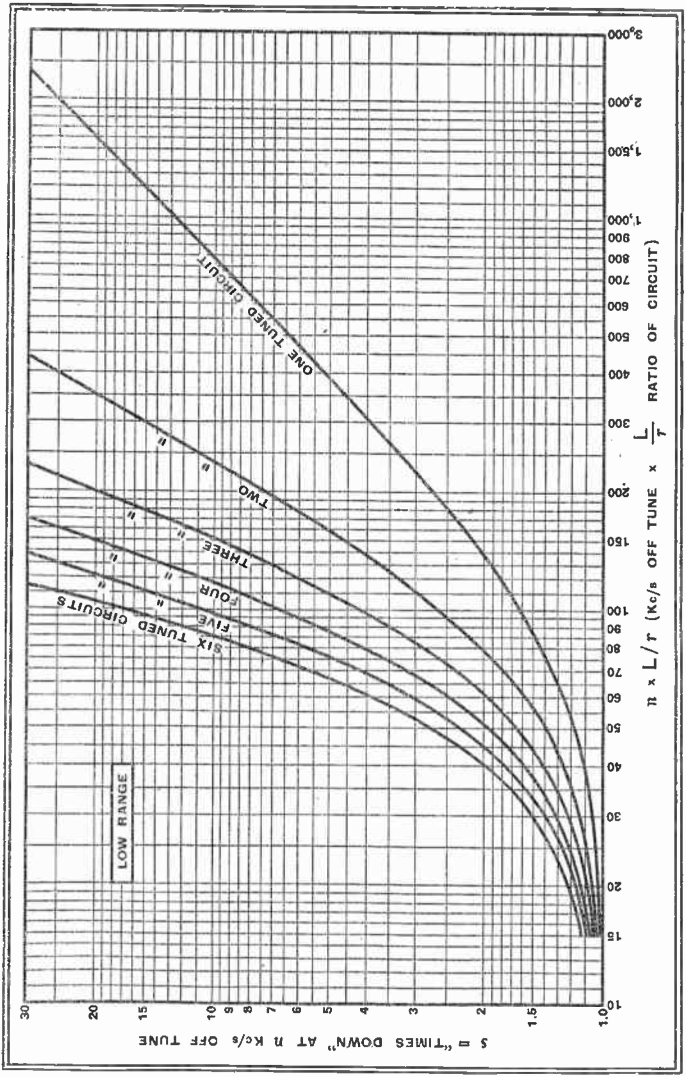


Fig. 115a: General data-curves showing relationship between $n \times L/r$ and s for 1 to 6 tuned circuits in cascade. Where value of s exceeds 10, use the continuation of these curves in Fig. 115b

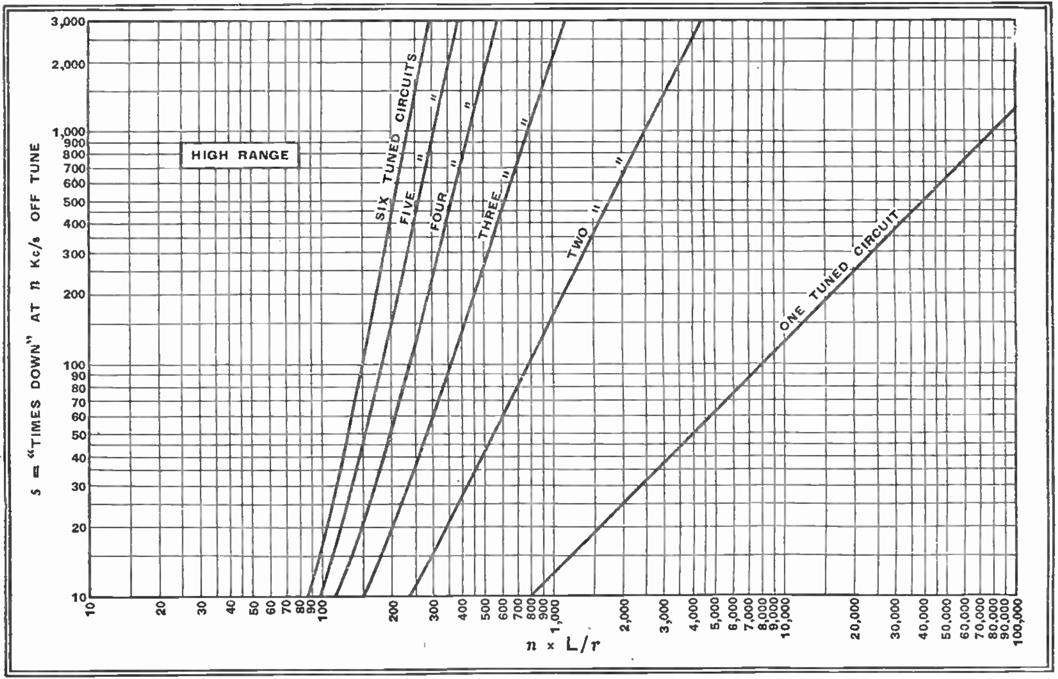


Fig. 115 b : Continuation of Fig. 115 a to higher values of $n \times L/r$ and s

At 5 kc/s off tune, $n \times L/r = 5 \times 8{\cdot}4 = 42$, corresponding on Fig. 115 to 2·1 times down. Thus, for the same discrimination against an unwanted carrier 18 kc/s removed from that required, six circuits give over 12 times as great a response to a 5-kilocycle sideband.

To make this point clearer, Fig. 116 shows the complete resonance curve, derived from

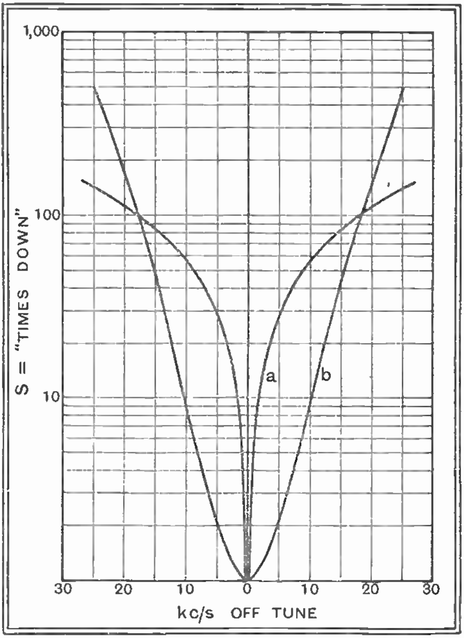


Fig. 116 : Resonance curves of one (*a*) and six (*b*) tuned circuits, chosen so as to give in each case 100 times reduction at 18 kc/s off tune. Note the enormous loss of sidebands in case *a*

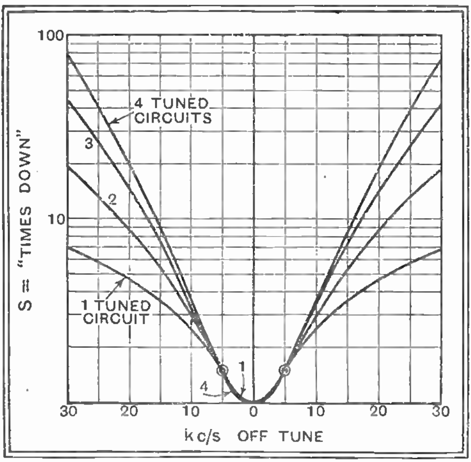
Fig. 115, for the two cases. The curves show very clearly that, although in both there is the same discrimination against a station 18 kc/s removed in frequency from that required, the single tuned circuit can only provide this selectivity at the cost off lopping off the sidebands of the desired transmission to a very drastic extent. The more rounded curve for six tuned circuits, though by no means perfect, offends very much less in this respect.

To reach so high a value of L/r as 450 it would be necessary to use a good deal of reaction, so that these two curves may be taken as illustrating, from a different angle, the dangers of trying to make reaction do too much. We have seen already how it destroys quality when used as a substitute for true amplification ; the curves of Fig. 116 emphasize that its use to provide selectivity that should be attained with additional tuned circuits brings just the same dire results in its train. These comments apply, of course, only to the excessive use of reaction ; in offsetting detector damping, and perhaps providing, in addition, a *little* extra selectivity or sensitivity it is invaluable, expecially in the less ambitious receiver.

128. Equal Quality

We have taken, perhaps, an extreme case in comparing the resonance curves of one and six tuned circuits. A more practical comparison is that shown in the four curves of Fig. 117. Here we can see the differences in selectivity obtained by using one, two, three, or four tuned circuits, the L/*r* values in each case being chosen to give $1\frac{1}{2}$ times down at 5 kc/s—that is, a reduction of 5,000-cycle notes to two-thirds of their correct voltage. This corresponds to a barely noticeable loss at this frequency.

Fig. 117 : Overall resonance curves of one, two, three and four tuned circuits, in each case chosen to give "equal quality", represented by the same response to a 5-kc/s sideband



For one tuned circuit we require that L/*r* = 18, which is by no means an outrageous value. The selectivity is poor, a station even three channels (27 kc/s) away being reduced only some six times. With two tuned circuits L/*r* for each comes out at 11, and a station 3 channels away is now reduced about 16 times. Adding a third circuit and reducing L/*r* to 8·8 still keeps the quality unchanged, but increases the selectivity to 32 times down at 27 kc/s. A fourth circuit increases this figure to 52.

129. Practical Coil Figures

It is simple enough, on paper, to talk about the choice of correct L/*r* ratios to provide the response-curves that we desire. In practice it is not always easy, or even possible, to achieve them. Experience shows that a coil designed to tune, with its variable condenser, over a range of wavelengths, always has a lower resistance at the longer wavelengths. For constant selectivity, one would, of course, require that the R.F. resistance should remain unchanged.

In the ordinary small coils used in the modern receiver, the ratio L/*r* is found to vary from about 5 or 6 at 1,500 kc/s (200 metres) up to about 20 at 550 kc/s (about 550 metres). If the coil has an iron-dust core and is wound with stranded wire in which the strands are insulated from one another ("Litzendraht") the 550-metre figure will probably rise to about 35, that for 200 metres remaining approximately unchanged. The design of a coil for lowest attainable resistance requires the choice of correct wire-thickness, and the thickness required depends on the precise wavelength for which the calculation is made. The figure given as representative for L/*r* can therefore be increased a little at either end of the waveband at the cost of a decrease at the other by designing the coil specifically for the wavelength it is desired to favour. But the only really useful method of improving the L/*r* ratio is by increase in size of coil; this, of course, is effective at all wavelengths.

130. Selectivity and Gain

It is an unfortunate fact that the less the selectivity changes over the wave-band, the less constant will be the gain. Gain depends, as we have seen (Sec. 113) on the dynamic resistance $R = \dfrac{L}{Cr}$ (Sec. 59); as we increase wavelength by increasing C, *r* diminishes and tends to hold constant the product C*r*, and with it the dynamic resistance, since L does not change. It is usually found that R has a maximum at about 240 metres, after which it falls steadily, till at 550 metres it is usually about half the maximum value.

If we really succeeded in keeping *r*, and hence L/*r*, constant from 200 to 550 metres, we should get constant selectivity accompanied by a steady drop in R which, at 550 metres, would have less than one-sixth of its value at 200. Conversely, constant R would give us marvellously constant gain, but to get it *r* would have to decrease in the same ratio that C increases, making L/*r* over six times as great at 550 metres as at 200.

Tuning by varying L, keeping C constant, could theoretically avoid this difficulty, for then constant R would also mean constant L/*r*. But *r* shows no particular inclination to be strictly proportional to L in any variable-inductance tuner that has so far appeared.

* 131. Long Waves

On the long-wave band, from some 800 metres up to 2,000, the coils generally used have an inductance of round about 2 millihenrys in conjunction with an L/*r* ratio varying from 30 to 50 over the band. On these wave-lengths higher figures can quite easily be attained, but they are hardly desirable on account of the severe loss of sidebands to which they give rise.

Consideration of the various figures that have been mentioned will make it clear that the ordinary set reduces the side-bands to a considerable extent, and yet suffers to some degree at least from insufficient selectivity. In spite of many attempts, the problem of making a satisfactory compromise between the conflicting claims of selectivity and quality is really not soluble in the case of the radio-frequency amplifier. A nearer approach to the desired results can be attained in a superheterodyne receiver, in connection with which we shall return to the question in Chapter 16.

CHAPTER 13

AUDIO-FREQUENCY AND OUTPUT STAGES

132. Resistance-Coupled Amplification

AUDIO-FREQUENCY amplification, by which is meant amplification of the signals after detection, is generally carried out in modern sets by some form of resistance amplifier. In Chapter 7 this method of amplification was discussed fairly fully, being taken as the type of amplification in general.

In Chapter 11 we found the method unsuitable for radio frequencies owing to the inevitable stray capacitances. In dealing with audio-frequencies these strays are naturally less harmful, but they may lead to a certain loss in gain at the highest notes, for which their reactance is of course least, if care is not taken in the choice of component values.

133. High Note Loss

It can be shown that high notes of frequency f are reduced to 70·7 per cent. of their correct voltage when $1/2\pi fC = R\ r_a/(R + r_a)$ (Fig. 118); that is to say, when the reactance of the stray capacitance is equal to the anode resistance of the valve in parallel with the load resistance. Any reduction in reactance (increase in capacitance) or increase in R or r_a leads to greater proportionate loss of high notes. It will be clear that where a high capacitance is inevitable (as in long screened leads, for example, or feeder lines to a distant amplifier) the choice of a valve of low anode resistance, with an external coupling resistance of low value, will ensure that loss of the higher frequencies is kept within reasonable bounds. If the frequency for which the equation given above is satisfied lies at 10,000 cycles or over, all will be well—and there will be a margin in hand to cover any underestimate of either capacitance or resistance.

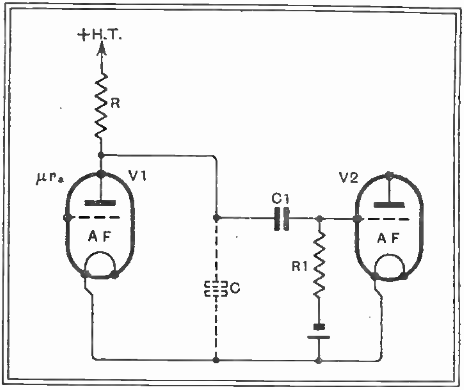


Fig. 118: Showing stray capacitances C in an L.F. stage. High notes of frequency f receive 70.7% of the amplification which they would have were it not for the presence of C, when $1/2\pi f C = Rr_a/(R + r_a)$

134. Low Note Loss

In Fig. 118 the grid condenser and leak, C_1 and R_1 form a potential divider across the source of amplified voltage (anode of V_1 to earth). Only the voltage appearing on R_1 reaches the grid of V_2, any dropped on C_1 being lost. For the lowest frequencies, at which its reactance is highest, there may be an appreciable wastage of signal on C_1; correct relative values must be chosen if this is to be avoided.

Low notes of frequency f are reduced to 70·7 per cent. of their full voltage when $1/2\pi f C_1 = R_1$; and, of course, a reduction in R_1 or in C_1 further increases the proportionate loss of low notes. A usual combination is $C_1 = 0{\cdot}01\ \mu F$, $R_1 = 0{\cdot}5\ M\Omega$, with which a note of frequency about 32 cycles is reduced to 70·7 per cent. Doubling either C_1 or R_1 will reduce this frequency to 16 cycles, but it is doubtful whether the resulting improvement in bass reproduction would be noticeable with the average loudspeaker.

135. Transformer Coupling

A transformer is often substituted for the resistance with the double aim of allowing a greater D.C. voltage to reach the anode of the A.F. amplifying valve (or detector) and of

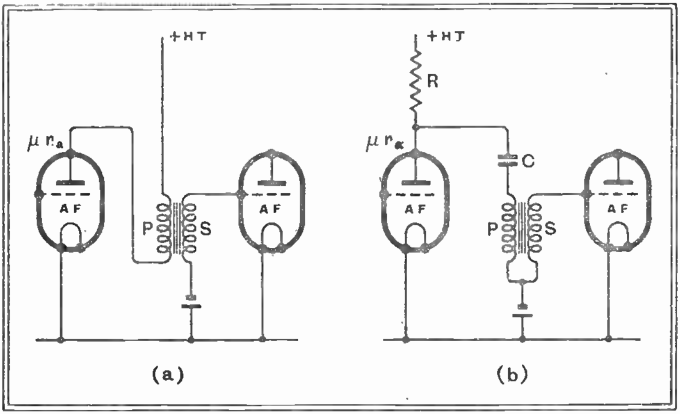


Fig. 119 : Transformer-coupled A.F. stages. In *a* the steady current of the first valve passes through the transformer primary P ; in *b* it is carried by R

obtaining extra gain by virtue of the step-up ratio of the transformer.

In Fig. 119 there are shown skeleton diagrams of a transformer-coupled stage. Since we desire to amplify signals of all frequencies to the same extent, the voltage developed across the primary in circuit *a* must be independent of frequency. The primary constitutes an inductive load, the reactance of which rises with frequency; to attain even amplification it follows, therefore, that the voltage across it must be substantially equal to μV_g at even the lowest frequency in which we are interested, since it will certainly rise to within a fraction of this figure at the highest. For this, the inductance of the primary must provide a reactance which, even at a low frequency, is high compared with the anode resistance of the valve. In the "equivalent anode circuit" of Fig. 120, the primary inductance L_p is in series with the anode resistance r_a of the valve, and receives $2\pi f L_p / \sqrt{r_a^2 + (2\pi f L_p)^2}$ of the generated voltage μV_g. If the primary reactance $2\pi f L_p$ is equal to r_a, the voltage across L_p will be 0·707 of what it is at a high frequency at which $2\pi f L_p$

considerably exceeds r_a. If we accept this condition as representing a tolerable drop in gain at the low frequencies we have at once a convenient design formula : $r_a = 2\pi f L_p$. For a given valve and transformer, this tells us the lowest frequency that is satisfactorily amplified; for a 10,000-Ω valve and a transformer for which $L_p = 100$ H, the drop to 70 per cent. of maximum amplification will occur at $f = r_a/2\pi L_p = 15{\cdot}9$ cycles. Evidently, with so good a transformer as this a valve of higher r_a, and hence higher μ, might be chosen. If we are content to set our limit at 50 cycles, then, with the same transformer : $r_a = 2\pi \times 50 \times 100 = 31{,}400$ ohms. This, therefore, is the highest permissible value of valve resistance. Or if r_a stays at 10,000 Ω, we can use a less bulky transformer, for which L_p is given by $L_p = r_a/2\pi f = 10{,}000/2\pi \times 50 = 31{\cdot}8$ henrys.

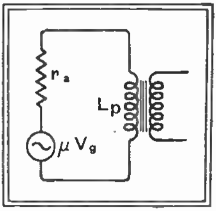


Fig. 120 : Equivalent anode circuit of a transformer-coupled stage

It is important to note that the necessary value for the primary inductance is that which holds in actual use, with the steady anode current of the valve passing through the winding. The permeability of the iron core, on which the inductance depends, falls off very severely when the magnetising force due to the current in the coil exceeds a certain amount. This is described as magnetic saturation. So a large initial anode current tends to prevent the core. from responding to the signal current, which has to superpose on it the varying magnetization from which the secondary derives its energizing voltage. In other words, the inductance is decreased below its "open-circuit" value by the steady current.

136. The Resistance-fed Transformer

This effect can be allowed for by making sure that the minimum value of L_p prescribed by the formula is reached even with the steady current passing through the winding, or alternatively by diverting the steady current through another path, as in Fig. 119 *b*. Most modern transformers have cores of high-permeability material (Mu-metal, etc.) which attain magnetic saturation with quite a small primary current. For these the "parallel" circuit shown at *b* is essential. The feed-condenser, if large enough, has negligible effect on the voltage across L_p at any frequency, but by cunning choice of a suitable value for C it may be made to maintain the bass response of a transformer at frequencies lower than that to which it would respond satisfactorily with a condenser of infinitely large capacitance. In effect C and P form a tuned circuit, tuning flatly on account of R and r_a which are virtually in parallel across it, by which the extreme bass can be maintained. Instructions for the choice of C, R, and r_a are generally given in the instruction-slip accompanying a transformer.

The matter of high-note response from a transformer is a complex one, depending partly on the stray capacitance across the transformer—which should evidently be kept at a minimum—and on a transformer characteristic (leakage inductance) not usually known to the ordinary user. Owing to the lack of available data on this point, no discussion of high-note response will be embarked on here.

137. The Output Valve

When amplified sufficiently, the signal is passed from the last valve in the set to the loud speaker, there to move a diaphragm which recreates, with more or less fidelity, the sound-waves from which the original modulation was derived. To agitate the diaphragm of a loud speaker *power* is required ; the output valve has therefore to be so chosen, and so worked, that the greatest possible amount of power is delivered to the loud speaker. To provide large power, high anode current and high anode voltage are required ; an output triode is therefore a valve of low anode resistance and may be rated to operate at voltages up to 400.

The properties of an output valve are deduced, in much

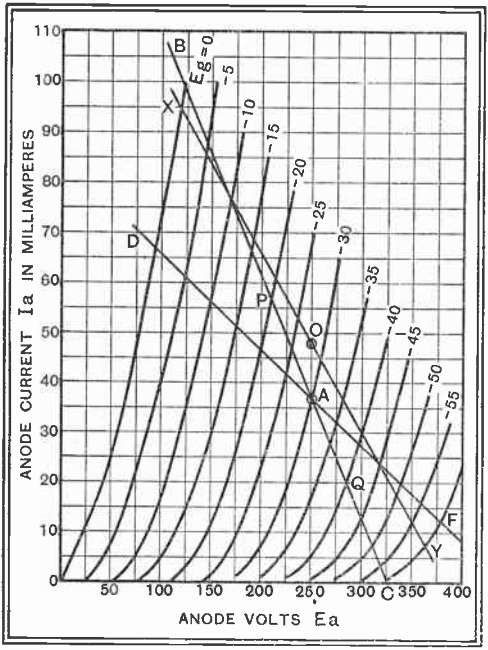


Fig. 121 : Curves of an output triode rated for 250 v. max. E_a. The load-lines shown are discussed in the text

the manner already discussed in Chapter 7, from load-lines drawn across the E_a — I_a curves. A set of such curves for an output triode are reproduced in Fig. 121. In discussing a resistance-coupled stage we saw that the load-line (Fig. 66) cuts the line $I_a = 0$ at the voltage of the anode battery, thereby indicating that the voltage at the anode of the valve could only rise to this value at zero anode current. In the case of an output valve the load consists of the windings of the speaker itself or of an output transformer, either of which has a comparatively low resistance. If, for the sake of simplicity, we regard this resistance as negligibly low, the voltage at the anode of the valve will be that of the anode battery itself, and the curves of the valve-plus-loudspeaker combination, if measured with direct current, will be those of Fig. 121. Let us suppose, then, that we decided to work the valve at the rated $E_a = 250$ v., and that we set the bias at — 30 v. This gives the working point A, for which $I_a = 37$ mA.

Even though the speaker offers no resistance to D.C., it will have quite a large impedance to signal currents ;

if we consider this impedance as purely resistive and as having the same value for all the frequencies in which we are interested, we can represent it by a load-line passing through A. Since the anode resistance of the particular valve illustrated is about 1,000 Ω, we will try a load-line of 2,000 Ω, on the grounds that the best load is usually greater than r_a. This line is shown at BAC.

138. Second-Harmonic Distortion

If we apply a signal of 10 volts peak the anode current will now swing between P and Q, or from 57 to 20 milliamps. The rise for the positive half-cycle is thus 20 mA, the fall for the negative half-cycle only 17 mA. This difference, clearly enough, will introduce distortion. Unless the grid-swing is restricted to uneconomically small dimensions, the distortion will not *entirely* vanish. We therefore have to set a more or less arbitrary limit to the amount of distortion we propose to permit ; that generally accepted allows distortion equivalent to the introduction of 5 per cent. of second harmonic. This is reached when the lengths AQ and AP stand in the ratio 9 to 11.

In the present case the grid-swing may be extended to about 15 volts each way, giving a change in I_a of + 30 and − 25 mA before this limit of distortion is reached. Corresponding to this total current-swing of 55 mA, there is a voltage-swing of 110 volts. The corresponding peak-values of signal current and signal voltage in the load are 55/2 and 110/2, and the R.M.S. values $55/2\sqrt{2}$ and $110/2\sqrt{2}$. The A.C. power delivered to the speaker is the product of these, or $(55 \times 110)/8 = 0{\cdot}756$ watt or 756 milliwatts.

139. Finding the Best Load

The restriction of the grid-swing made necessary by the early attainment of the 5 per cent. distortion limit indicates that the load has been wrongly chosen. Going through the same process of drawing load-line, investigating permissible grid-swing before the distortion-limit is reached, and calculating from the current and voltage swings the power delivered to the speaker, enables us to find the power that can be delivered into each of a series of loads of different impedance. The results are given as a curve in Fig. 122.

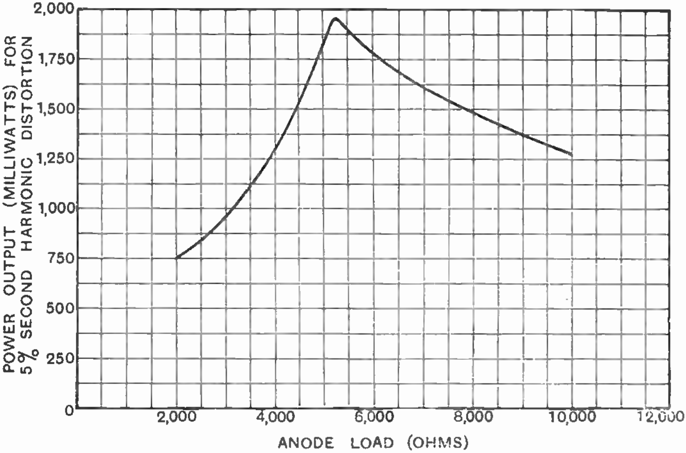


Fig. 122 : Relation between anode load and available power, allowing 5 % second-harmonic distortion, for the valve of Fig. 121 worked at point A

The *optimum load*, being that into which the greatest power can be delivered, is evidently about 5,200 Ω—the corresponding load-line is drawn at DAF on Fig. 121. To achieve this power the grid requires a signal that swings it from 0 to − 60 v., giving a swing in anode current from $12\frac{1}{2}$ to 67 mA. The two excursions from A are exactly in the ratio 9 to 11, showing that distortion equivalent to the introduction of 5 per cent. second harmonic has just been reached. The power available for the loud speaker is now

$$\frac{(67 - 12\frac{1}{2}) \times (378 - 94)}{8} = \frac{54\frac{1}{2} \times 284}{8} = 1935 \text{ mW}$$

It will be remembered that the choice of A as the working-point was purely arbitrary—it is quite possible that some other point would give greater power. Still keeping to $E_a = 250$ V., which, being the highest voltage for which the valve is rated, will quite certainly give the greatest output,* other points can be investigated in the same

* The power output given by a valve is related to the anode voltage applied thus : Power is proportional to $(E_a)^{5/2}$

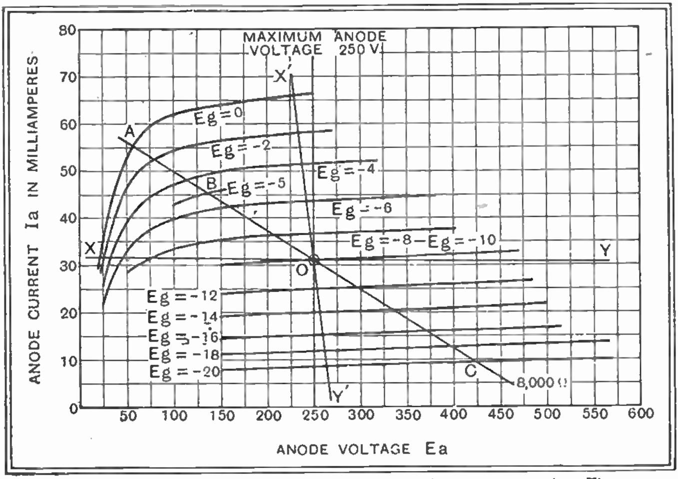


Fig. 123 : Curves of a typical indirectly heated output pentode. The load-line ABOC represents a usual load. Curves taken at E_S = 200

manner as A, and then, by comparing the outputs given by thc bcst load for each point, we can finally pick the best possible working-point and load. For the valve of Fig. 121, this is given by $E_a = 250$, $I_a = 48$, $R = 2{,}930\ \Omega$. For this, the available power is 2,670 mW, as can be deduced from the final load line XOY.

In general, the user of a valve is not compelled to go through this elaborate examination of valve-curves, for the makers' recommendations as to anode voltage and current, grid bias and optimum load are set forth in the instruction-slip accompanying each valve. The user has only to do as he is told.

This is just as well, because actually, although there is a middle range of frequencies over which a loudspeaker presents an approximately resistive load, at the extreme frequencies the reactance predominates, and this makes the matter too complex to deal with here. The resistive load line does, all the same, give a useful guide.

In the matter of providing the optimum load the user is rather at sea ; he can do no more than ask the maker of his chosen loudspeaker to supply it with a transformer suited to the valve he proposes to use. The ratio of the transformer, as reference to Section 48 will show, should be $\sqrt{\frac{R_p}{R_s}}$ where R_p and R_s are respectively the required load and the mean impedance of the speech-coil.

The ordinary tetrode is not suitable as an output valve owing to the distortion that would occur when the signal swung the voltage at the anode close to or below that of the screen (Sec. 121). But a pentode, or a special form of tetrode in which secondary emission has been suppressed to give it the typical pentode characteristic, can be used as an output valve.

140. The Output Pentode

Compared with the triode, the pentode offers the two advantages of being more *efficient*, in the sense that a greater proportion of the power drawn by its anode circuit from the H.T. supply is converted into A.C. power for operating the speaker, and of being more *sensitive*, in that a volt of signal applied to its grid produces a larger output. For these two reasons the pentode has largely supplanted the triode as output valve for sets where cost is a prime consideration.

Screened and output pentodes differ in minor points, but not in principle. In the latter, since screening is no longer vital, grid and anode are both taken to pins in the base. High output is obtained by designing the valve to operate with a screen voltage little, if at all, below that at the anode.

In Fig. 123 are reproduced the curves of a typical indirectly heated output pentode; their similarity to the usable portion of the curves of a tetrode will at once be evident. We see again the high anode resistance (curves nearly horizontal) typical of valves using a screening-grid between control-grid and plate.

141. Loading the Pentode

In the case of a pentode, the usual triode rule that the anode load should be greater than the anode resistance of the valve does not hold. At the working point O (E_a = 250 v., E_g = − 10 v., I_a = 31 mA) the resistance of the valve is some 125,000 Ω (change in I_a of 2 mA brought about by change in E_a of some 250 v.) ; XOY is a load-line representing 250,000 Ω drawn through O. Towards X it cuts the curves for $E_g = -8$ to $E_g = 0$ in very rapid succession, while towards Y it looks as though it will never reach the curves for $E_g = -12$ to $E_g = -20$. With a load such as this, the application of a signal swinging the grid from 0 to -20 would very evidently result in the most appalling distortion, together with the development of amazingly high audio-frequency voltages at the anode. (At what value of E_a does the line XOY cut the curve $E_g = -20$?)

If we were to fly to the other extreme and draw a load line (X′OY′) representing a very low load, distortion would again result, owing to the line now cutting the curves for high bias in very rapid succession, while the intercepts with the low-bias curves are widely spaced. Since these two types of distortion, for high and low loads respectively, occur at opposite ends of the total grid-swing, it is fairly evident that some intermediate load is going to be found best.

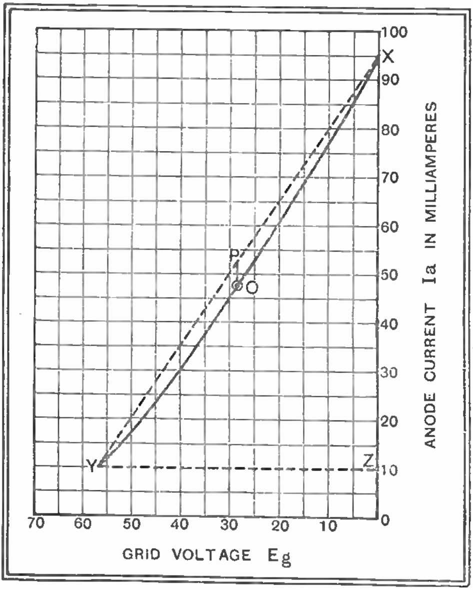
We are led to the same conclusion if we consider the power developed (still for the grid-swing 0 to -20 v.) in the two loads. XOY offers high voltages and negligible current, while X′OY′ provides high current but negligible voltage. To get both voltage and current reasonably large an intermediate value of load is clearly required.

Let us investigate an 8,000-Ω load, which experience suggests as a possible load for a pentode. This is indicated by the line ABOC. The power delivered to this load when a signal swings the grid from $E_g = 0$ to $E_g = -20$ can be obtained, as with a triode, from the voltages and currents at the points A and C ; it is

$$\frac{(56{\cdot}2 - 9{\cdot}2)}{2\sqrt{2}} \times \frac{(424 - 56)}{2\sqrt{2}} = \frac{47 \times 368}{8} = 2{,}160 \text{ mW.}$$

142. Harmonic Distortion and the Pentode

How about distortion? With the triode, as we have seen, the distortion anticipated is second-harmonic dis-



tortion, and we accepted the convention that the permissible limit of this is 5 per cent. With the pentode we haveto take into account distortion equivalent to the introduction of both second and third harmonics of the original signal.

In Fig. 124 is plotted the dynamic characteristic of a triode working under conditions of 5 per cent. second harmonic ; the data for this are taken from the load-line XOY of Fig. 121. To show up the non-linearity of the curve, a straight line joins its extremities ; the divergence between the current at the actual working point O and that shown, for the same bias, on the straight line, is the measure of the second-harmonic distortion. Calling the currents at X and Y respectively I_{max} and I_{min}, that at P is midway between the two, or $\frac{1}{2}$ (I_{max} + I_{min}). The difference between this and I_o, the actual current at O, divided by the total current swing (I_{max} — I_{min}), gives the proportion of second harmonic, requiring only to be multiplied by 100 to give the percentage. The formula for calculation is thus: Percentage second harmonic $= \dfrac{\frac{1}{2}(I_{max} + I_{min}) - I_o}{I_{max} - I_{min}} \times 100.$

Fig. 124 : Dynamic curve of output triode giving 5% second-harmonic distortion, and, in dotted line, ideal characteristic for no distortion. Percentage second harmonic $= \frac{PO}{XZ} \times 100$

To introduce third harmonic, as with the pentode, the curve must bend *both ways*, as in Fig. 125, which shows the dynamic curve of a valve introducing about 12 per cent. third harmonic, but zero second. Freedom from second harmonic is shown by the fact that O now lies on the straight line joining A and C, but it will be seen that the curve lies below the line between C and O, and above it between O and A. This particular type of divergence from linearity usually implies third harmonic. It can be numerically estimated in a similar way to second-harmonic distortion, using now only *half* the curve. It is approximately found from the difference between the actual current at B and the current at D (which, being on the straight line, is the mean between the currents at O and A), this difference being divided by the total change in current in passing from O to A.

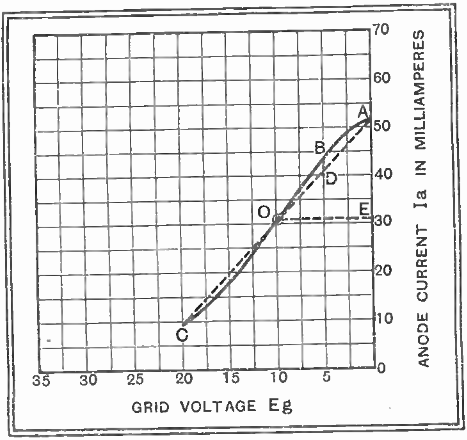


Fig. 125 : Dynamic curve of pentode with 10,000-ohm load. Ideal characteristic giving zero second and third harmonics is shown dotted. Percentage third harmonic = $\frac{BD}{AE} \times 67$ approximately

143. Relation Between Load and Distortion

By drawing a number of load-lines across the curves of Fig. 123 and calculating second- and third-harmonic distortion for each, the results summarized in the curves of Figs. 126 and 127 have been obtained. The difference between the two sets of data is that in making the calculations for Fig. 126 it was assumed that the signal had a peak voltage of 10 v., thus swinging the grid between zero and -20 v., whereas in Fig. 127 the calculations have been made for an 8-volt signal, swinging the grid from -2 to -18 v. only. As might be expected, the distortion is much less for the restricted input.

In both cases the second-harmonic distortion is high for a low load, but drops away to zero as the load is increased. This is the load for which the dynamic characteristic has the form shown in Fig. 125. Still higher loads reintroduce second-harmonic distortion, which then rises rapidly with increasing load. Third-harmonic distortion, as the curves show, increases steadily with increasing load, as does the power delivered to the speaker. It is from a number of curves such as these, calculated not for one but for several alternative working points, that the final operating data for a pentode are determined by its designer.

The "high-slope" pentode, at present much used in certain types of set, only differs from the standard type by requiring a much smaller signal-voltage. A typical valve of this class will yield about 2,500 milliwatts in

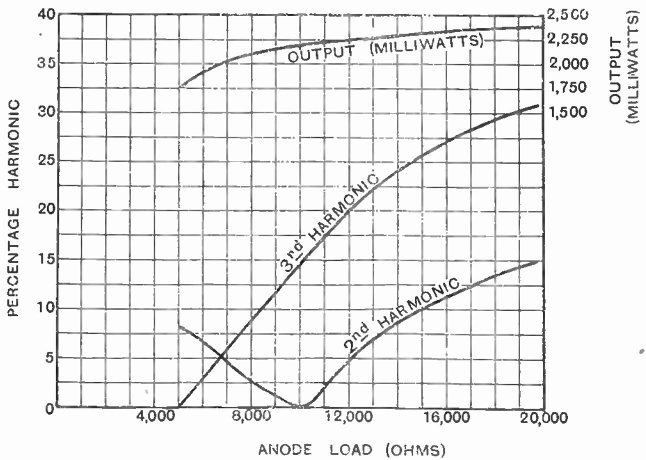


Fig. 126: Output and second and third harmonic distortion for pentode of Fig. 123. Working-point O; input signal 10 v. peak

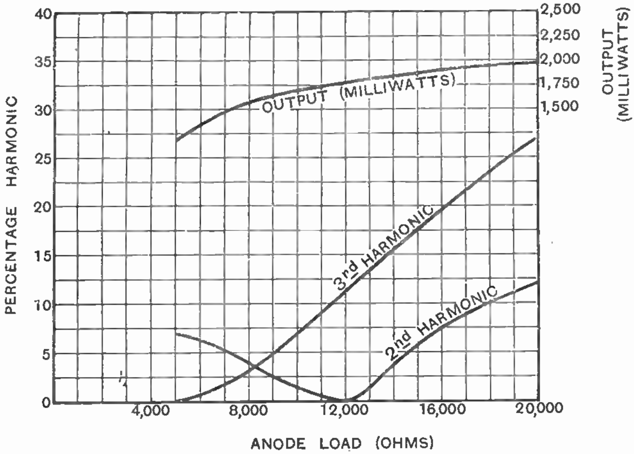


Fig. 127 : Output and second and third harmonics for pentode of Fig. 123. Working-point O ; input signal 8 v. peak

return for a signal of 3 v. peak, instead of the 10 v. needed by the standard pentode we have been discussing.

144. Negative Feedback

The advantages of a pentode, which are high gain and high output power on moderate anode voltages, are to some extent offset by the too-ready development of third-harmonic distortion.

It is generally agreed that distortion containing a large percentage of third harmonic, or even a small percentage of higher harmonics, is more objectionable than that associated with second harmonic.

If a reduction in the gain of the valve can be tolerated, it is possible to decrease very considerably the proportion of harmonics in the output without decreasing the power available. This is done by feeding back into the grid-circuit a small proportion of the amplified voltage present at the anode.

This can be done in any one of several ways, but it is necessary, in order to maintain the high input impedance of the valve, that the voltage fed back into the grid-circuit should be inserted in series, and not in parallel, with the original signal voltage. The circuit of Fig. 128 is a very suitable one for the purpose, the voltage fed back being that developed across R_2, the lower member of the potential divider across the output. C, of capacitance about 1 μF., serves simply to isolate R_1 and R_2 from the D.C. voltage at the anode of the valve. To avoid appreciable loss of output power, R_1 and R_2 together should have about ten times the load resistance, and it is usually desirable to make R_2 about one-fifth to one-eighth of R_1, thus feeding back from one-sixth to one-ninth of the output voltage.

The effect of this feedback is to reduce the gain to about one-fifth of its normal value, so that the preceding stage must deliver five times the usual signal-voltage to the pentode grid. As the gain-reduction occurs through reducing this input voltage by an opposing voltage fed

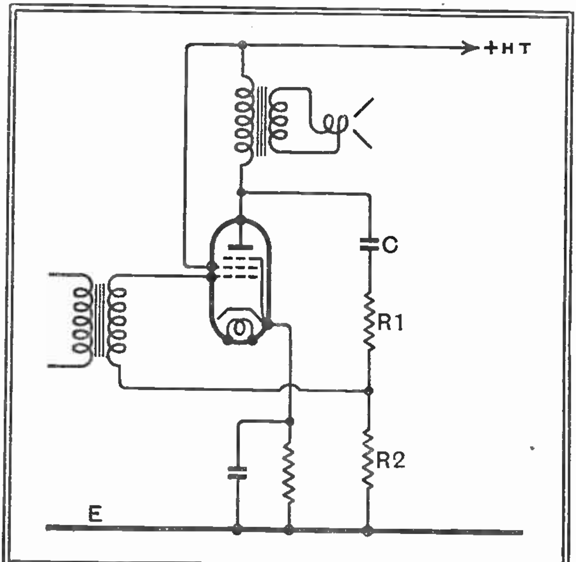


Fig. 128 : Circuit for negative feedback. R_1 and R_2 form a potential divider across the output, the voltage developed across R_2 being fed back into the grid-circuit in series with the transformer secondary.

back, the pentode does not actually handle any increased signal, and so requires only its normal bias.

The harmonic-content of the output voltage can be shown to be reduced in about the same proportion as the gain of the stage, so that under conditions where a pentode would normally give a gain of 30 times, accept a signal of 5 volts, and deliver 2 watts of power with 10 per cent. harmonic distortion, the addition of negative feed-back might reduce the gain to six times, and make it necessary to supply a signal of 25 volts. This would result in 2 watts of power with only about 2 per cent. harmonic distortion.

Although the load required by the valve is unchanged by the introduction of negative feedback, its apparent resistance is enormously reduced. The new value of this is

$$\frac{r_a}{1 + \dfrac{r_a + \mu R_2}{R_1 + R_2}}$$

and with the normal values used for the circuit this amounts approximately to dividing the anode resistance of the valve by $\dfrac{\mu R_2}{R_1 + R_2}$. As μ for an output pentode may be of the order of 600, and the ratio $R_2/(R_1 + R_2)$ will be about one-seventh, the anode resistance of the valve when feed-back is used is not far from one-hundredth of its normal value.

The practical value of this is in dealing with the fact that loudspeakers inevitably resonate at certain frequencies, giving excessive prominence to reproduction of them, as well as causing " ringing " or prolongation of the sound beyond that present in the original performance. By shunting the loudspeaker with a low resistance, such resonances can be flattened out or " damped " just like any tuned circuit. A triode output valve is such a low resistance, but a pentode is not, and an appreciable part of the poor quality of reproduction associated with pentodes can be traced to speaker resonances. But by means of negative feedback applied to a pentode, its resistance effective for damping out loudspeaker resonances can be made even lower than in a triode without feedback.

With the addition of negative feedback, and at the cost of no more than a reduction in gain, a pentode gives as great freedom from speaker resonances as a triode and gives less than the triode's harmonic distortion, while retaining the high power-output and moderate bias of the pentode.

Care must be taken that the increased "drive" demanded by the grid of the output valve when negative feedback is used does not overload any preceding stage. To avoid this risk, and at the same time extend the benefits of feedback, the output may be fed back over more than one stage, perhaps from the secondary of the output transformer. A wide variety of circuits have been worked out to suit circumstances.

145. Valves in Parallel and in Push-Pull

If more power is wanted than can be provided by a single output valve, two (or more) may be used. By simply adding a second valve in parallel with the first, connecting grid to grid and anode to anode, the swings of voltage at the anode are left unchanged, but the current

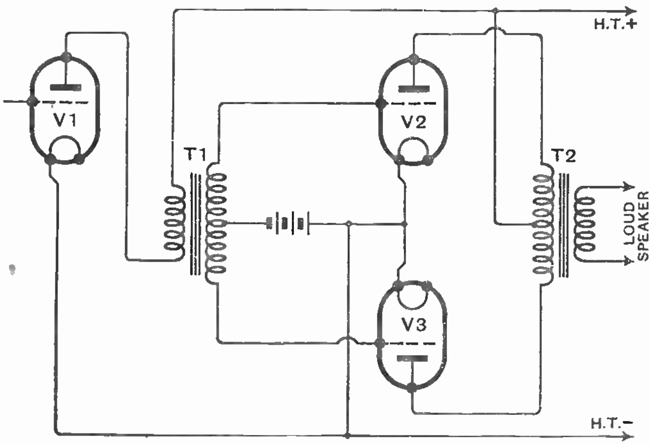


Fig. 129 : Two output valves, V_2 and V_3, in push-pull. The same circuit also applies to Q.P.P. and "Class B", the differences being only in the operating voltages and choice of valves

swings are doubled. So, therefore, is the power, while the load resistance needed for two valves is half that needed for one. The performance of the whole output stage can be deduced from the $E_a - I_a$ curves of one of the valves merely by multiplying the figures on the anode-current scale by the number of valves it is proposed to use.

Alternatively, the valves may be connected in *push-pull*, as shown in Fig. 129. Here the output valves are fed from a transformer T_1, in which the mid-point, instead of one end, of the secondary is earthed. At an instant when, with the normal connection, the "live" end of the secondary would be at + 20 v., the other (earthed) end being zero potential, the centre-point of the winding would be at + 10 v. With the push-pull arrangement this centre-point is brought to earth potential, the two ends, therefore being respectively + 10 and − 10 v. Thus each valve receives half the available voltage, the two halves always being in opposite phase.

The resulting out-of-phase anode currents, which would cancel one another if passed in the same direction through a transformer, are made to add by causing them to flow through separate halves of a centre-tapped primary, as shown at T_2 in Fig. 129. The voltage induced into the secondary, and hence the current flowing in the loud speaker, is due to the combined currents of the two valves.

This mode of connection has several advantages over the more obvious parallel arrangement. These are :—

(1) The steady anode currents, since they pass in opposite directions through their respective primaries, cancel one another so far as saturation of the core of the transformer is concerned. A smaller transformer can, therefore, be used for two valves in push-pull than for the same two valves in parallel.

(2) Signals fed through the common H.T. connection cancel ; valves in push-pull are, therefore, unable to feed magnified signals into the H.T. line of a set, and so cannot give rise to undesired feed-back. Conversely, disturbances on the H.T. line (hum, etc.) cancel in the two valves.

(3) *Second*-harmonic distortion produced by either valve is cancelled by equal and opposite distortion from the

other. Two *triodes* in push-pull will, therefore, give a greater undistorted output than they would if connected in parallel.

Third-harmonic distortion does not cancel in this way. Pentodes, whose output is limited by third harmonics, (see Fig. 126), consequently give no greater output in push-pull than in parallel. Advantages (1) and (2), however, apply to pentodes as much as to triodes.

146. Q.P.P. and Class B

If valves, whether triodes or pentodes, are over-biased, the distortion arising is mainly second-harmonic distortion. With two valves in push-pull, this type of distortion will automatically vanish. Two valves in push-pull may, therefore, be given so large a bias that their anode current is reduced practically to zero, making them behave, on receiving a signal, as though they were anode-bend detectors. So biased, the valves of Fig. 129 will each amplify only during the moments when its grid is made more positive by the applied signal, during which instants the anode current rises in proportion to the signal voltage applied. If the valves would normally be biased to —10 v., each would then require a 20-volt total grid swing making the total swing on the transformer secondary 40 volts. Both valves would then amplify at every instant, and the standing anode current might perhaps be 20 mA per valve, remaining almost unchanged on the application of the signal.

Now, suppose each valve biased to — 20 v., and the signal doubled. The no-signal anode current might now be only 3 mA per valve, the two valves giving alternate kicks up to 40 mA when the full signal is applied. The *average* anode current, even when delivering full output, is less than in a normally biased push-pull stage; while if the applied signal is well below the maximum that the valves can handle, the average current, made up now of alternate kicks up to perhaps 6 mA, is quite small. Since, on a musical programme, the full output of the valves is only called for at brief and infrequent moments, this trick of overbiasing results in a very large overall saving of anode current without curtailing the available output. In mains sets, where anode current costs practically nothing, this device is seldom used; in battery sets, where anode current costs perhaps twenty to one hundred times as much, it has found wide application. The system is called *quiescent push-pull*, commonly abbreviated to Q.P.P., and specially designed output valves are offered by several makers. Owing to the need for doubling the input signal, the less sensitive triode is seldom used, each half of the Q.P.P. output valve being usually a pentode.

Another quiescent output scheme designed to economize anode current is found in the *Class "B"* output stage, which again uses the basic circuit of Fig. 129. In this case the two output valves (usually combined in one bulb) are high-resistance triodes taking, as in Q.P.P., only a small anode current except when a signal is applied. The bias used is at most small, with the result that *the grids are swung heavily positive* by the signal. Grid current inevitably flows, thereby consuming audio-frequency power; the preceding valve must therefore be so chosen that it can deliver this power without overloading, while the transformer feeding the Class "B" valve must be a properly designed "*driver*" transformer of the correct ratio and of low D.C. resistance. By removal of the no-grid-current limitation large powers can be obtained from a Class "B" output stage at the cost of a remarkably low average anode current.

As in the case of output stages of other types, many details of the performance of push-pull, Q.P.P., or Class "B" output stages can be obtained by careful study of the appropriate $E_a - I_a$ curves; but the design of Class "B," especially, is very complex, and likely to give poor quality and disappointing results generally unless many factors are taken into proper consideration.

147. The Loudspeaker

Whatever output stage is used, the amplified currents in the anode circuit of the last valve eventually reach the loudspeaker, the duty of which, as we have already seen, is to convert the audio-frequency currents into corresponding air-waves. More strictly expressed, it has to convert the audio-frequency electrical power supplied to it into acoustic power at the same frequency. As in every case where electrical energy is converted directly into mechanical energy, this is done by taking advantage of the magnetic field set up by the current.

Fig. 130 shows the cross-section of an energized speaker, in which the magnet is provided by passing a current through the winding A. Through the centre of this winding runs an iron rod B, the purpose of which is to guide the lines of magnetic force due to the current. This it does because the permeability of iron to the lines is very high, and they therefore pass through the iron in preference to the air in much the same way that an electric current passes through a copper wire, and not through the air around it. The analogy is not complete, because the air does carry some lines ; there is no " insulator " for magnetic lines, but only materials of very high " resistance ". The high permeability of iron as compared with air results in the iron core enhancing the intensity of the field as well as directing it, much as a conductor of low resistance will carry a larger current between two points of different electrical potential than will one of high resistance.

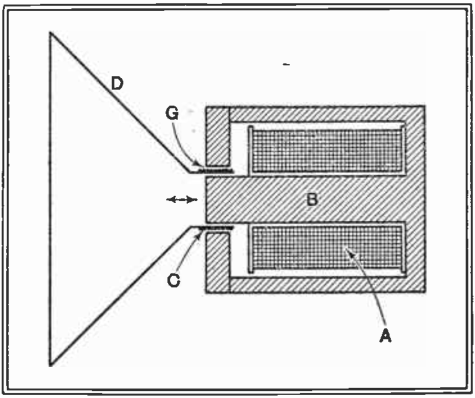


Fig. 130 : Cross-section of a moving-coil loudspeaker

The outer shell of the cylindrical magnet is also of iron, so that except for the small circular gap at G there is a complete iron circuit. The lines are thus guided round the iron and are all made to complete their path by jumping the gap, in which there is, in consequence, an extremely concentrated magnetic field.

In this gap is suspended the coil of wire C, wound on a former firmly attached to the paper diaphragm D. If we lead a current through C the coil will tend to move along the gap, driving D towards or away from the face of the magnet according to the direction of the current.

In the anode circuit of the output valve of a set receiving a tuning-note there is flowing an alternating current of frequency equal to that of the note. If C is connected in that anode circuit, it is driven in and out, as suggested by the arrow in Fig. 130, at the frequency of the current, and so the diaphragm, moving with it against the resistance of the air, converts into acoustic energy the power supplied by the valve. It thus sets up an air-wave conveying to the ear, at a loudness depending on the power in C, a note at the frequency of the current.

If the signal has the enormously more complex waveform of a piece of orchestral music, the movements of the coil, and hence of the diaphragm, still follow it—or would, in a perfect speaker—so faithfully reproducing that music.

It will be evident that at an instant when the diaphragm in Fig. 130 is moving to the left, there will be compressed air in front of it and rarefied air behind it. If the period of one cycle of movement of the diaphragm is long compared with the time in which the resulting air wave can travel round its edge from front to back, these pressures will equalize and no sound will be sent out. To prevent this loss, evidently worst at the lowest notes, the loud-speaker is always mounted so that it " speaks " through a hole in a *baffle*. This consists of a piece of wood, flat or in the form of a cabinet, designed to lengthen the air-path from front to back of the diaphragm and so to ensure that the bass is adequately radiated.

CHAPTER 14

DESIGNING A SIMPLE SET

148. The Specification

WE have covered, fairly fully, all the essential points necessary for the full comprehension of every part of an ordinary receiver not of superheterodyne type, but the references are scattered about over the twelve preceeding chapters. Before going on to consider the peculiar properties of the superheterodyne, it is proposed to devote a short chapter to the practical discussion of the design of a typical simple set, with the idea of making a kind of summary of the ground already covered. In discussing the various points that arise we shall have to take for granted conclusions already reached. In order to help the reader to look up any points about which he may be doubtful, numbers in brackets refer him to the section in which fuller elucidation may be found.

We will suppose that we have been asked to design a set which will have an average sensitivity of about one millivolt. By this is meant that if a carrier-voltage of this magnitude, modulated to a depth of 30 per cent. (78) is applied to the aerial terminal, the overall magnification of the set will be such that the " standard output " of 50 milliwatts of modulation-frequency power (138) will be delivered to the loudspeaker. The selectivity of the set is to be that associated with three tuned circuits—since their L/r ratio is bound to vary widely over the wave-range covered (129) no numerical specification of selectivity is practicable. The whole is to be driven by batteries, and, for the sake of economy in upkeep, is to consume a maximum of 10 milliamps. in the anode circuits.

149. The Outlines of the Circuit

The first points to be settled are the type of output stage to be used, the kind of detector we shall choose, and whether the three tuned circuits shall be associated with one or with two radio-frequency amplifying valves. These points are inter-related and involve also the limitation in total anode current already imposed.

This latter limitation immediately suggests the choice of a quiescent output stage (Q.P.P. or Class " B ") (146), but also implies that a small-size H.T. battery is likely to be used. Now small batteries generally fail, except when new, to hand out the large instantaneous currents (146) demanded by quiescent output stages, and by so failing introduce very evident distortion. We will therefore play for safety and choose as output valve a pentode, on the grounds that it makes more noise per milliamp. than does a triode (140).

A battery pentode, if of the high-resistance type, takes about 5 mA at 120 v., in return for which it will deliver some 250 to 300 mW before overloading. This, though small, is an acceptable output for a set of the type contemplated. Allowing another milliamp. for the screen of the pentode, 6 of our available 10 mA are already accounted for.

With only three tuned circuits in the set it is quite certain that occasions will arise when the selectivity will not be adequate for separating the station required from others on neighbouring frequencies (125). In order that selectivity can be enhanced when desired, reaction will have to be available to the user (101). The use of fairly flatly-tuned circuits with adjustable reaction as an auxiliary will enable the inevitable selectivity-quality compromise (102 ; 126) to be readjusted by the user as he tunes from station to station.

For providing reaction the diode detector (63 ; 84) is obviously useless. The anode-bend detector is not good from this point of view either, because it depends for its results on being operated on a part of its characteristic curve where the slope is very slight (92), and hence delivers comparatively little power to the anode circuit for reaction. We shall therefore choose a grid detector (88 ; 100).

Either a screened pentode or a triode may be successfully used for this purpose, the former giving much the higher amplification. To set against this advantage it has so high an anode resistance (about 0·5 M Ω) that the use of a transformer to couple it to the output pentode is out of the question if we have any respect at all for our low notes (135). Shunting the transformer by a resistance (136) would limit the high-note gain to that available for low notes, but in so doing the gain would be reduced to about that of a simple triode. If we try to use resistance coupling, the voltage at the anode will be found to be seriously restricted by the voltage-drop in the resistance, and detector overload (88) will set an uncomfortably low limit to the available output, especially with deep modulation (78). To provide our output pentode with the signal (approximately 3 v. peak) that it needs to develop full output, and at the same time to make reaction behave satisfactorily, it will be safest to choose a triode detector followed by a transformer of step-up ratio not less than one to three.

True, we shall now have serious input damping (99), which we could have avoided by choosing a screened valve, but reaction will take care of this (101). Unless a little reaction is used this input damping will make tuning rather flat, and sensitivity perhaps a shade disappointing. But by attention to tuned-circuit design this effect can be considerably reduced, as we shall shortly see.

To avoid all risk of overloading, even on low modulation, we shall hardly be safe if we allow the detector less than about 1 to 1½ mA of anode current—which, with the 6 mA of the output valves, leaves us 2½ to 3 mA for the R.F. side of the set. This is about the current of a single screened valve, but by biasing back we could keep the total current of *two* valves within this limit, and still have more gain than one valve could yield. What gain do

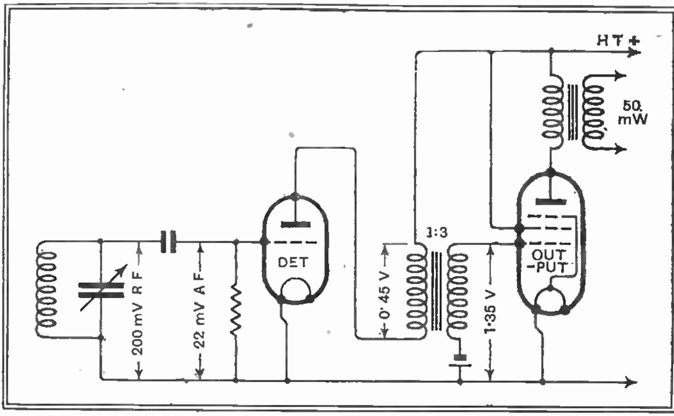


Fig. 131 : Approximate evaluation of voltages on pentode and detector for 50 mW output

we need? To find this we must work back from the output valve, as in Fig. 131.

150. Checking the Sensitivity

The pentode gives 250 mW for a 3-v. peak signal; 50 mW, therefore, for a signal of $3/\sqrt{5} = 1{\cdot}35$ v. peak across the secondary of T. Across the primary, assuming a 1 : 3 ratio, we shall require 0·45 v. If the detector valve has $r_a = 20{,}000\ \Omega$, $\mu = 24$, under operating conditions, we can reckon on an audio-frequency gain of getting on for 20 times from grid to anode, so that we shall require a rectified signal, inside the grid condenser, of about 0·022 v. or 22 mV.

For so low an input as this implies, detector efficiency will be very low (104), and, over-emphasizing this inefficiency so as to be on the safe side, we might reckon that 200 mV of carrier-voltage, modulated at 30 per cent., will be needed to produce a rectified signal of this magnitude.

This tells us that for a sensitivity of one millivolt we must have a radio-frequency gain of about 200 times between aerial terminal and detector grid. The gain given by one valve, ignoring detector-damping, will be about 60 times (113; but those figures referred to a *mains* valve, which has a higher g_m) from grid of R.F. valve to grid of detector, so that we shall need some 3 to 4 mV at the first valve's grid. Across the second of two coupled circuits, the voltage is usually about four to eight times that actually applied to the aerial terminal, owing to the step-up effect of the tuned circuits (51) ; we see, therefore, that 1 mV on the aerial terminal will comfortably give us the required 50 mW output with only a single R.F. valve, provided that, as assumed, reaction is used to an extent just sufficient to offset detector damping (101). We shall certainly not need a second R.F. valve ; in fact, if we were to use one, the sensitivity of the set would be too high for its selectivity. By this is meant that the additional stations brought in by the extra sensitivity, being necessarily those which give only weak signals at the aerial, would all be liable to serious interference from stronger ones. Unless it were added simply with a view of making up for the deficiencies of a tiny aerial, the extra sensitivity would therefore be of no value in practice.

151. The Circuit Completed

Our set, then, will be arranged thus: two tuned

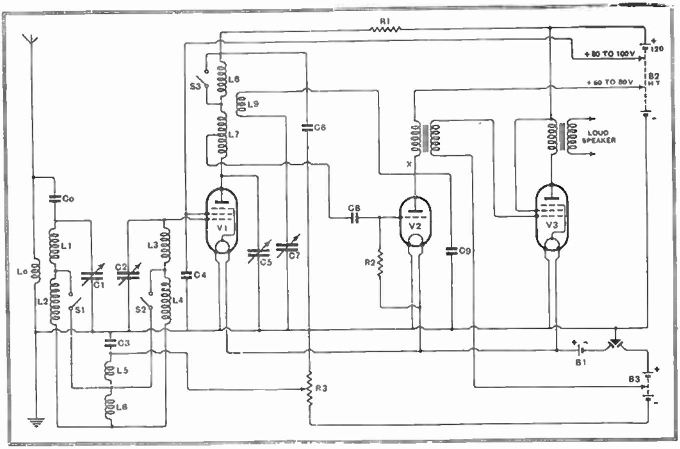


Fig. 132 : Complete circuit of three-valve set to conform with specification laid down in Section 148

circuits, R.F. valve, tuned circuit with reaction, grid detector, transformer, output pentode. Such a bald skeleton description as this does not prescribe an exact circuit ; a dozen designers would produce a dozen circuits all differing from one another in minor ways. One of the many possible variations on the theme is shown in Fig. 132, where the complete receiver, including wave-band switching, is shown.

Careful inspection of this rather elaborate diagram will show that it really consists of an assembly of separate circuits, each of which, regarded individually, is by now perfectly familiar. With but one or two unimportant exceptions, every separate circuit has been discussed somewhere or other in past pages. Dissection of the diagram is best performed by tracing grid, anode and screen circuits right through, starting at the electrode in question and continuing, through H.T. or bias battery, until the cathode of the valve is reached. Observe that sometimes the same components can be common to two circuits—for example, the tuned circuit $C_5L_7L_8C_6$ is included both in the anode circuit of V_1 and in the grid circuit of V_2.

Some small points in the circuit may be puzzling at first sight, even though their meaning could be seen by arguing from basic principles. The coupling of aerial to first tuned circuit is done by the combination of the primary winding L_0 and the condenser C_0, of capacitance about 20 $\mu\mu$F. The two together, if suitably dimensioned, can be made to give more or less constant step-up at all wavelengths on the lower (medium-wave) band. On long waves, S_1, S_2 and S_3 are open so that the tuning inductances in use are $L_1 + L_2$, $L_3 + L_4$, and $L_7 + L_8$. One section of each composite coil is shorted out for medium-wave reception.

Energy is transferred from the first tuned circuit to the second by making the coil $L_5 + L_6$ (on medium waves, L_5 only) common to both circuits (compare 109), so that the voltage developed across it by the current in the first circuit acts as driving voltage for the second. L_5 will need to be about 3 μH, while L_5 and L_6 together will be about 30 μH. The condenser C_3 is inserted to close the circuit for R.F. currents while allowing a variable bias, taken from the potentiometer R_3 connected across the bias battery B_3, to be applied to the grid of the variable-mu screened pentode V_1 to control its amplification (120).

The tuning condenser C_5 goes from anode to earth instead of directly across its coil L_7L_8 in view of the fact that C_1, C_2, and C_5 will normally be in the form of a three-gang condenser, with rotors on a common spindle. The tuned circuit is completed through the non-inductive condenser C_6, which, in order to maintain the ganging of the set, should have the same capacitance as C_3. Each may be 0·25 μF. or over ; much less would begin to reduce the tuning-range appreciably (38).

Since R.F. currents flow in the anode circuit of the detector, which is completed through the H.T. battery B_2, any R.F. voltage developed across this will be conveyed to the anode of V_1, and so to the grid of V_2. The resistance R_1, of some 5,000 Ω, serves as protection against instability from this cause.

Damping imposed by the detector on the tuned circuit is decreased, if only for medium waves, by connecting the detector grid to a tap on L_7. If the tap is at the centre of the coil, damping will be reduced to one-quarter(96). The reaction-coil L_9 is coupled to both L_7 and L_8, and the current through it is controlled by the variable condenser C_7. The inductance of the reaction coil must be such that C_7 does not tune it to any wavelength within the tuning range of the receiver, or reaction control will be difficult. The increase in sensitivity and selectivity (101) produced by applying reaction will also be felt in the circuit $L_3L_4C_2$, owing to a certain amount of energy feeding back through the screened valve and by way of stray couplings (107 ; 114).

As shown, the circuit does not include a radio-frequency choke in the anode circuit of the detector, the primary of the A.F. transformer T_1 serving as substitute. This attempted economy may lead to difficulty in obtaining proper reaction effects. Alternatively, by allowing R.F. currents to stray into the output valve, and then back, via loudspeaker leads, to the aerial side of the set, it may lead to hooting and grunting noises when receiving a signal, especially when much reaction is being used. In such cases an R.F. choke must be inserted at X, making sure that the anode by-pass condenser of the detector (100) is still directly connected to the anode.

As shown, the set requires three positive connections to the H.T. battery. This enables the technically-minded user to adjust the voltages at detector anode and pentode screen either for maximum sensitivity or for economy of current. In a commercially-built set, to be handled by non-technical users, it would be better to provide a resistor of fixed value in each of the movable leads and to take them all to maximum H.T. voltage.

It is hoped that this chapter has given the reader a glimpse of the way in which all the various matters discussed in earlier parts have to be brought together when considering the design of a set, and of the process by which a concrete design emerges from a brief specification of intended performance. Any reader who may be taking this book really seriously may like to complete the design here only begun; by a sufficiently close study of earlier chapters he could find a suitable value for every component in the set, after which, adding some data from a valve catalogue, he could work out, at least approximately, the overall sensitivity, selectivity, and fidelity of the receiver at a number of different wavelengths.

CHAPTER 15

THE SUPERHETERODYNE AND ITS FREQUENCY-CHANGER

152. The Need for Selectivity

SOME idea of the problem of separating one station from another when they are spaced only 9 kc/s apart, without losing the sidebands of the wanted station, will have been gained from Chapter 12. The problem is intensified by the demand of the listener to be able to hear a station undisturbed even when its next-door neighbour (in frequency) is many times more powerful than itself. It is complicated still more by the fact that selectivity varies as the tuning condenser is adjusted.

In the usual "straight" set, as discussed in the last chapter, pre-detector amplification is carried out at the frequency of the signal. So long as we have only two or three circuits to be retuned every time we pass from one station to another, this system is convenient enough, but if we were to demand selectivity of so high an order that ten tuned circuits were needed to provide it, the set would become impossibly cumbersome.

153. The Principle of the Superhet

When high selectivity in conjunction with simplicity of control is required, the supersonic heterodyne receiver (conveniently known as the "superhet.") is the only possible type of set. In Fig. 133 is given a schematic diagram of a superhet., in which the various parts of the set are shown as labelled boxes. Of their contents we shall speak later.

The signal received from the aerial is first put through a stage of *pre-selection*, containing tuned circuits enough to ensure that signals of wavelengths far removed from that of the station required shall not pass farther into the set. This box may or may not contain a stage of ordinary radio-frequency amplification of the type with which we are now familiar.

The next stage, the *frequency-changer*, operates upon the signal in such a way as to produce a carrier of a new frequency, this new carrier still carrying the modulation of the original carrier. In most cases the new carrier has a frequency lower than that of the original signal, though it is always *supersonic*, or higher than any frequency within the audible range. It is, in consequence, usually referred to as the *intermediate frequency*, commonly abbreviated to

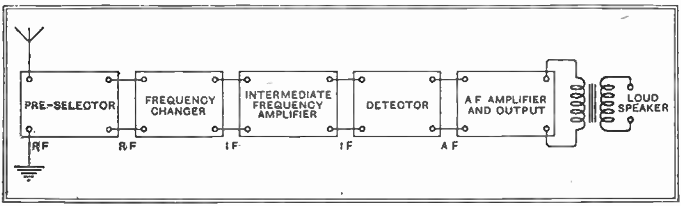


Fig. 133 : Schematic diagram of supersonic heterodyne receiver ("superhet.") showing the functions of the various parts and the changes in the signal in passing through it

I.F. At this new frequency it undergoes further, and often considerable, amplification in the third box of Fig. 133, after which it is passed to detector and A.F. amplifier in the ordinary way. Pre-selection, frequency-changing, and I.F. amplification thus take the place of the R.F. amplifier and associated tuned circuits of an ordinary set.

Whatever may have been the frequency of the received signal, it always emerges from the frequency-changer at the one fixed intermediate frequency, this being determined by the designer of the set. To perform this conversion, the frequency-changer subtracts from (or adds to) the signal frequency whatever frequency is necessary in order to bring it to the I.F. As the signal frequency varies according to the station it comes from, whereas the I.F. is fixed, it follows that the frequency to be subtracted must be varied to suit the station it is desired to receive, and therefore the frequency-changer must be tuned to subtract the right frequency. If it is so tuned that a 1,000 kc/s signal has its frequency converted to 465 kc/s (by subtracting 535 kc/s) ; and signals at 991 and 1,009 kc/s, due to stations working on *adjacent channels* in the frequency band, are also present, the frequency-changer converts them to 456 and 474 kc/s respectively. It follows that if the I.F. amplifier is accurately and selectively tuned to 465 kc/s, these two signals will not pass through it, and hence will not reach detector, A.F. amplifier, or loud-speaker.

Adjacent-channel selectivity, or selectivity aimed at removing stations on frequencies closely bordering on that of the desired station, can therefore be provided entirely by design of the I.F. amplifier without reference to any other part of the set.

Since the I.F. amplifier is tuned to the one fixed frequency, it becomes practicable to include in it just as many tuned circuits as are needed to provide the selectivity we require ; they have only to be tuned once, when the set is first made. Nor is this the only advantage of operating on a fixed frequency ; by careful and finicky adjustment we can shape the overall resonance-curve to give us any desired compromise between selectivity and sideband response with the comforting knowledge that this compromise will hold unchanged for every station received. Further, its constancy permits of judicious faking of the A.F. amplifier to strengthen high notes if we find that we cannot get the selectivity we desire without undue cutting of side-bands in the I.F. tuned circuits.

Although the highest usable adjacent-channel selectivity can be provided in the I.F. amplifier, tuning is still required in the pre-selector stage. This is so because the characteristics of the frequency-changer are such that stations *on certain wavelengths widely removed* from that of the station required can set up in it a carrier of the intermediate frequency, so causing interference with the station to which the set is intended to be tuned. It is the duty of the pre-selector to eliminate these outlying frequencies before they can reach the frequency-changer, leaving the task of providing adjacent-channel selectivity to the I.F. amplifier.

154. The Frequency-Changer

The function to be performed by the frequency-changer, as we have just seen, is to convert all the signal frequencies into new frequencies, of which only those coming from the desired station must be allowed to get through the I.F. amplifier. Suppose a sample of the carrier wave coming from the desired station in one hundred-thousandth of a second to be represented by Fig. 134*a*. As it contains 10 cycles, the frequency is 1000 kc/s. If another signal generated in and by the frequency-changer itself is as shown at *b*, its frequency is 1,450 kc/s. Now add the two together. The result is shown at *c*, and can be seen to be a 1,450 kc/s wave varying in amplitude at a rate of 450 kc/s. But it is only a variation (or modulation), at 450 kc/s, of a signal of higher frequency. *No signal of 450 kc/s is present*, for each rise at that frequency

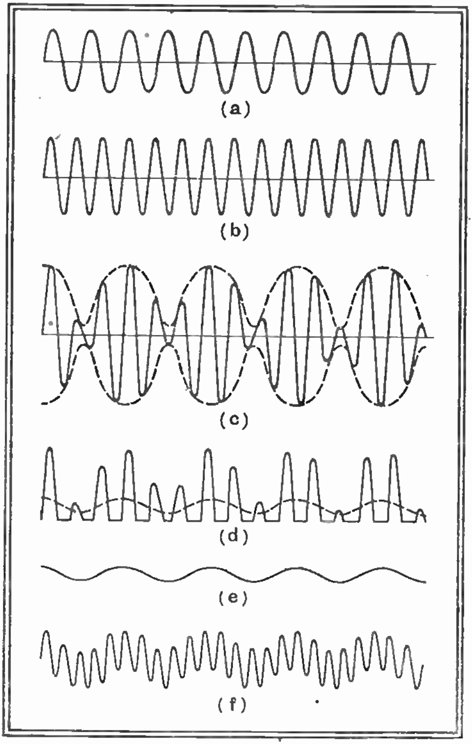


Fig. 134: A sample of incoming carrier wave, of ten millionths of a second duration, is represented at *a*. To the same scale *b* is the local oscillator signal. When the two are added (in a "first detector") the result is c, which must be rectified (d) to yield a difference frequency (e). But if *a* and *b* are *multiplied* together (*f*), the difference frequency comes directly

above the centre line is neutralised by an equal and opposite fall below the line. The average result, when applied to a circuit tuned to 450 kc/s, is nil.

We have been up against this difficulty already—see Fig. 75—and the solution is the same now as then; rectify it. By eliminating all the negative half-cycles (*d*) the smoothed-out or averaged result of all the positive half-cycles is a 450 kc/s signal—the difference between the frequencies illustrated at *a* and *b*—shown dotted at *d*. This mixture of frequencies is passed to the I.F. tuned circuits, which reject all except the 450 kc/s, shown now alone at *e*. The unwanted frequencies are so far away from that to which the I.F. circuits are tuned that they are soon eliminated.

So far, the result is only a 450 kc/s carrier wave of constant amplitude, corresponding to the constant amplitude of the 1000 kc/s carrier wave (*a*) received from the aerial. It is not difficult to see, however, that if *a* is modulated the modulations are repeated in *e*; so long, at least, as *a* does not exceed *b* in amplitude. Looking at the modulation of *a* as the addition of sideband frequencies (Sec. 103), these sidebands when added to *b* and rectified give rise to frequencies that are sidebands of *e*. Suppose that our 1 kc/s tuning note is broadcast: in addition to the 1,000 kc/s at *a* there are 999 kc/s and 1,001 kc/s. When added to 1,450 kc/s the 450 kc/s beat itself beats at 1 kc/s, and when rectified there are present 449 kc/s and 451 kc/s. As the I.F. amplifier is designed to cover a band of several kc/s each side of 450, these sidebands are amplified along with the carrier wave; and so far as the following detector is concerned the combination might equally well have been due to a station transmitting on a carrier frequency of 450 kc/s.

The early types of frequency-changer, not commonly used now, worked on the principle just described. In some, the functions of mixer and rectifier (or, as it was generally called, *first detector*) were allocated to one valve and that of oscillator to another; in others, a single valve did everything.

155. A Two-Valve Frequency-Changer

An example of the former type is shown in Fig. 135. V_2 is the oscillator which, in essence, is an arrangement in which reaction is pressed so far that continuous oscillation results (Secs. 101 and 108). The resistance R_4 serves in lieu of an R.F. choke to divert the radio-frequency anode current through the reaction-coil L_3 besides being useful in limiting the average anode current of the valve. Further help in this direction is supplied by grid rectification of the oscillation, which biases V_2 negatively (Sec. 84). The

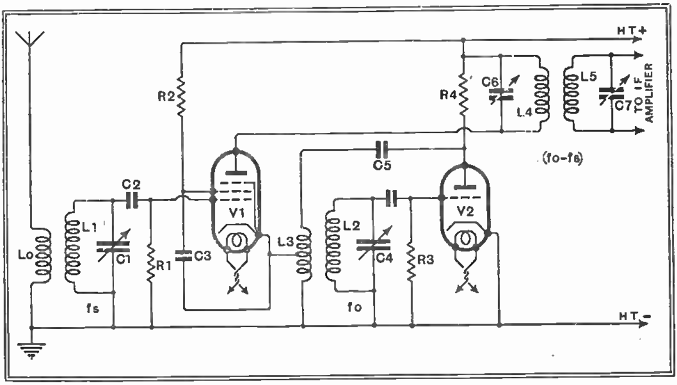


Fig. 135 : Circuit of simple two-valve frequency-changer. The signal-frequency *fs* is applied to the grid of V_1, and the oscillator V_2 is tuned to *fo*. Currents at the intermediate frequency (*fo* – *fs*) appear in the anode circuit of V_1

frequency of the oscillation is that to which the tunable circuit L_2C_4 is adjusted; for convenience of reference we will call this, the oscillator frequency, *fo*.

The signal, of frequency *fs*, is collected from the aerial or other source by the tuned circuit L_1C_1, and applied to the grid of V_1, the screened pentode used as first detector. The grid-condenser C_2 is not included for purposes of signal-rectification, but to enable the valve to set itself, by grid-current, at its correct working point. The cathode of V_1 is taken to a tapping on the reaction coil L_3, thereby including that part of L_3 that lies between tap and earth in the grid circuit of the valve. (Remember that the grid circuit includes everything between grid and cathode.)

The amplitude of the oscillation thus applied to the grid of V_1 will require to be about 10 to 15 v. peak in a circuit of this kind; suitable choice of tapping point on L_3 ensures a correct voltage. Like the oscillator, V_1 will bias itself back until the applied oscillation just, and only just, runs the grid into grid current. Assuming a 10-volt peak oscillation at this point (represented by *b* in Fig. 134), the bias of V_1 is being swung, at the frequency of the oscillation, from zero to -20 and back again.

The negative peaks therefore occur in a region of the valve's characteristics at which anode current is almost if not entirely cut off. The incoming signal, which relatively is far smaller than suggested by Fig. 134*a*, is meanwhile causing both positive and negative peaks to fluctuate at the intermediate frequency; though far less violently than shown at *c*. As the negative peaks are cut off, only the positive ones are effective in varying the anode current; and so we get *d*, from which the I.F. amplifier selects *e*.

156. Modern Frequency-Changers

The early types of frequency-changer, especially those in which one valve was made to do everything, fell short of the ideal requirements :

1. Maximum I.F. output for a given R.F. input.
2. Absence of radiation from oscillator.
3. Frequency stability of oscillator.
4. Ability to oscillate freely at very high frequencies.
5. Absence of "pulling" between oscillator and pre-selector circuits.
6. Application of volume control, and especially A.V.C. (Chapter 17).

All of these requirements became much more difficult to fulfil as the demand arose for reception of very high frequencies (short waves). One of the most tricky to deal with is 5. The I.F. in vogue during this early period in the development of the superhet was 110 kc/s. This represents a difference between received signal and local oscillation frequencies of 55 per cent. at 200 kc/s (long wave) or 11 per cent. at 1,000 kc/s (medium wave), which causes no great difficulty; but at 20 Mc/s (15 metres, short wave) the difference is only about 0·5 per cent. Any slight coupling due, say, to stray capacitance, between the oscillator and signal preselector tuning circuits tends to cause the latter to be pulled into step with the former, and to provoke other undesirable effects. As the products of both tuning circuits must be applied to one valve in order to combine them, the interelectrode capacitance must be reduced to a very small figure indeed if an undesirable amount of coupling is not to result; and even if it is eliminated by suitable screening there are other more subtle forms of coupling that give trouble. The situation was relieved by a general adoption of an I.F. in the region of 450-470 kc/s, increasing the disparity between the two frequencies.

As this alone did not solve all the problems, intensive work by valve designers has resulted in an almost bewildering variety of frequency-changers, some with separate oscillators but most combining all functions in one bulb. It is not possible in a limited space to deal with all of these, but one outstanding feature is a departure from the principle outlined in Sec. 154 (known as the *additive* principle because the incoming signal and the local oscillation are added together and then rectified in order to yield the desired difference frequency). Nearly all modern frequency-changers are of the *multiplicative* type, in which no rectification is needed, the difference frequency being produced directly. Look again at Fig. 134: curve *c* was formed by adding the heights above or below the centre lines of *a* and *b*. If these heights are multiplied instead of added, remembering that two negative heights when multiplied give a positive height, the result is curve *f*. The difference frequency, 450 kc/s, is here already without rectifying; and so is the sum frequency, 2,450 kc/s, which could quite possibly be used as the I.F. instead of 450 (there are as many crests in *f* as in *a* and *b* together).

But how does a valve multiply? It is easy enough to add two signals in a valve by feeding them in series between grid and cathode so that the net grid-cathode voltage is the sum of the two.

Sec. 113 showed that the amplified output of a valve of the high anode resistance type is practically equal to the input voltage multiplied by g_mR. R is the dynamic resistance of the output tuned circuit, which in a superhet

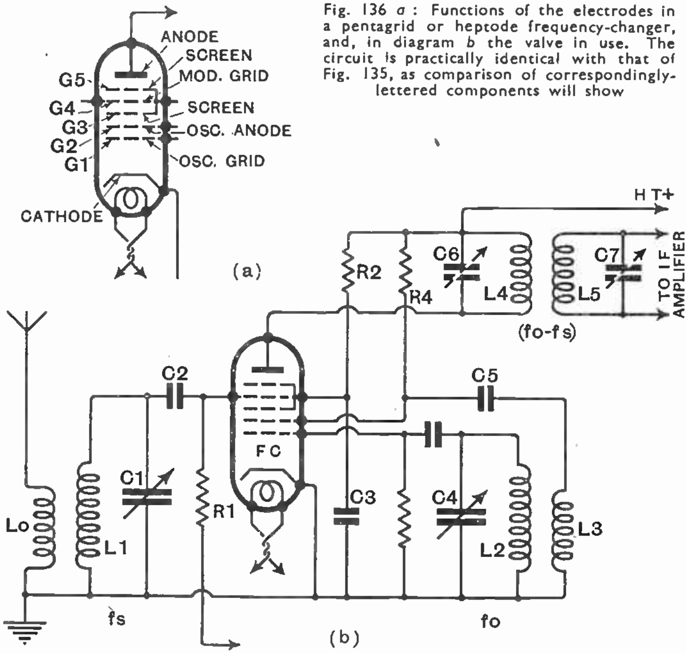


Fig. 136 *a* : Functions of the electrodes in a pentagrid or heptode frequency-changer, and, in diagram *b* the valve in use. The circuit is practically identical with that of Fig. 135, as comparison of correspondingly-lettered components will show

is the first I.F. transformer, so is fixed. If g_m, the mutual conductance, can be made proportional to the local oscillation voltage, then the output is proportional to the input signal voltage multiplied by the oscillation voltage at every instant.

The first successful valve of this type was the heptode or pentagrid, shown in Fig. 136.

As its name implies, the valve has five grids, the uses of which are shown on the diagram. G_1 and G_2 form the grid and anode of a triode oscillator, the circuit of which, as Fig. 136 *b* shows, in no way differs from that of V_2 in Fig. 135. G_4 and G_5 serve as control grid and screen of a screened tetrode performing the functions of V_1 in Fig. 135. The additional grid G_3, connected within the valve to G_5, serves to screen the modulator grid from the oscillator, and so prevents G_4 from biasing itself back, as does V_1 in the two-valve circuit. This valve, therefore, remains responsive to control of amplification by variation of bias ; G is consequently given variable-mu characteristics, and the controlling bias is fed to it through the resistance R_1.

The almost exact identity of the two frequency-changing circuits is emphasized by the fact that exactly the same components are used in both ; for convenience, they have been identically lettered in the two diagrams. The parallel can be made even closer by replacing the pentagrid with an *octode*, for in this valve there is yet another grid between G_5 and the anode, thus converting the tetrode outer portion of the pentagrid into a screened pentode.

The sole real difference between the two lies in the method of arranging that the oscillation shall vary the mutual conductance of the screened valve that deals with the signal.

Every electron that reaches the modulator (made up of G_4, G_5, and the anode) has to pass *through the oscillator* (G_1 and G_2) on its way.

The slope of the modulator is low when the oscillator grid is strongly negative, and high when its potential is zero or slightly negative. When oscillations are present on G_1, this grid will bias itself back until only the extreme positive peaks cause grid current to flow ; the total excursion of the grid will therefore be from approximately zero to double the peak voltage of the oscillation. It is therefore evident that we have drastic variations of modulator slope at the frequency *fo* of the local oscillations generated by the triode portion of the valve. Since the incoming signal, at frequency *fs*, is applied to the grid of the modulator, we have a system in which the amplitude of the original signal is in effect multiplied by that of the local oscillator, leading to the production of sum and difference frequencies in the anode circuit. Of these, that desired is picked out by the tuned circuit L_4C_6, and passed, through a second tuned circuit L_5C_7 to the I.F. amplifier valve.

Actually, as the slope of a valve is not a thing that can have negative values, the waveform shown in Fig. 134*f* is somewhat different in practice, the difference consisting in the presence of the original signal frequency along with the others, but that is also rejected by the I.F. amplifier.

Although there is screening within the heptode between the oscillator and preselector tuning circuits, ordinary capacitance is not the only form of coupling possible. The large fluctuating mob of electrons controlled by the oscillator section constitutes a *space charge* (Sec. 62), which induces a current at oscillator frequency in the preselector circuits. This effect becomes serious at the higher frequencies and, together with other still more obscure phenomena, causes the heptode to be largely ineffective at over 20 Mc/s. Some of the later types of octode have, by skilled design, been rendered useful at these very high frequencies ; but the most popular frequency-changer at the present time is the *triode-hexode*. Here we have virtually two valves, only the cathode being common to both—in point of fact, a separate triode and hexode are used in some sets—and the noteworthy feature is that the signal control grid comes first, and the really large movements of electrons, produced by the local oscillation, occur beyond, where they can exert very little undesirable influence on the preselector circuits.

Fig. 137 shows the connections, and again the lettering of the components is the same. A cathode resistor provides a small amount of initial bias for the signal, or modulator, grid. The grid of the oscillator section is internally connected to the oscillator injection grid (No. 3 in the hexode), which is screened from both signal grid and anode by screen grids. With some valves it is preferable for the oscillator anode to be tuned, rather than the grid. It is possible to give the triode section a higher mutual conductance than in the heptode, enabling oscillation to be readily obtained at higher frequencies in spite of the various influences that tend to increase losses at those

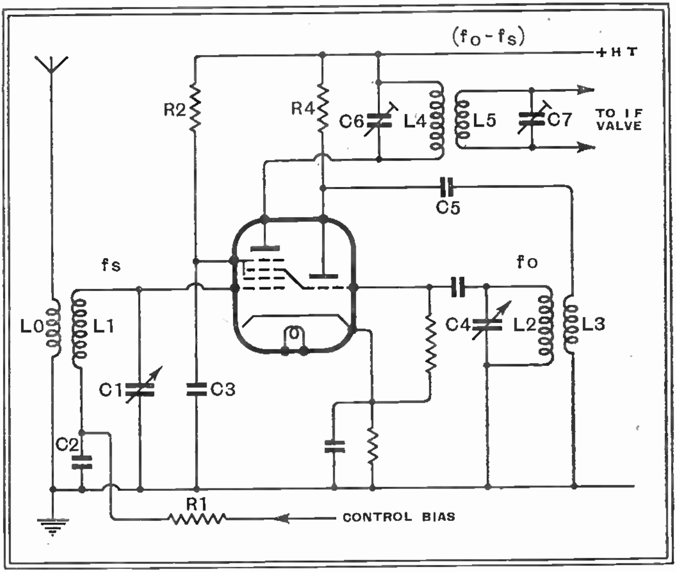


Fig. 137 : Typical triode-hexode frequency-changer circuit, lettered to correspond with Figs. 135 and 136*b*

frequencies ; and the other requirements specified at the beginning of this Section are also generally better met by the triode-hexode frequency-changer.

157. Conversion Conductance

In the ordinary amplifying valve the mutual conductance is expressed in terms of milliamps of signal-current in the anode circuit per volt of signal applied to the grid. The same rating can be applied to a frequency-changing valve, but it is not of much help in receiver design. In this particular case we are interested in milliamps of current *at intermediate frequency* per volt of signal (radio-frequency) on the grid. This is known as the *conversion conductance* of the valve, and quite evidently depends on the efficiency of conversion as well as on the amplifying abilities of the modulator.

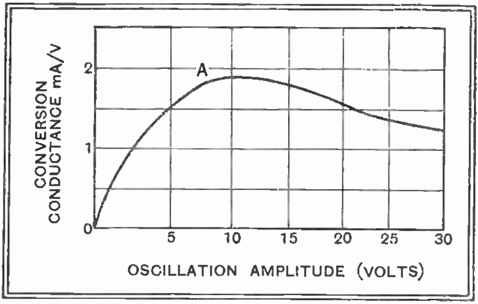


Fig. 138: Type of relationship between oscillation amplitude and conversion conductance of frequency-changer. The oscillation amplitude is not critical so long as it exceeds a certain value in the neighbourhood of A

Looking at Fig. 134, it can be seen that the peak fluctuations in the amplitude of *b* due to *a*, indicated in *c* by the dotted lines, are equal in amplitude to *a*, on each side of *c*. The fluctuation in *mean* amplitude, shown by the dotted line in *d*, which represents the I.F. output, is less. With full-wave rectification, it would be $\frac{2}{\pi}$, or 63 per cent. of the peak value. But half-wave rectification leaves gaps between each half-cycle, and this figure must be halved, giving $\frac{1}{\pi}$ or 37 per cent. Assuming perfect rectification, then, the conversion conductance with additive frequency-changing is a little over one-third of the mutual conductance. If the oscillation amplitude is insufficient, rectification is imperfect, and conversion conductance falls off, reaching zero, of course, when the oscillation amplitude is zero. On the other hand, if the oscillation amplitude is excessive, only the extreme peaks will pass anode current, there will be bigger gaps between them, and the mean value and consequently the conversion conductance will fall, though not so rapidly. There is therefore an optimum oscillation amplitude for every valve.

In Fig. 134*f* the amplitude of the I.F. component is half that of a, so the maximum conversion conductance is half the mutual conductance in the multiplicative type of frequency-changer, and therefore rather better than in an additive type having the same mutual conductance. The condition for optimum oscillation amplitude is that it shall be just sufficient to vary the mutual conductance from maximum to zero.

Fig. 138 illustrates this for a heptode, and it is evident that there is less danger of losing gain by too powerful than by too weak an oscillation. It is usual, therefore, to arrange that at no part of the wave-band to be covered by the set shall the oscillation amplitude fall below a value corresponding to a point at or near A on the curve. This is done by adjustment of turns on the reaction coil (L_3, Figs. 135, 136, and 137), after which the oscillator can be left to look after itself.

158. Ganging the Oscillator

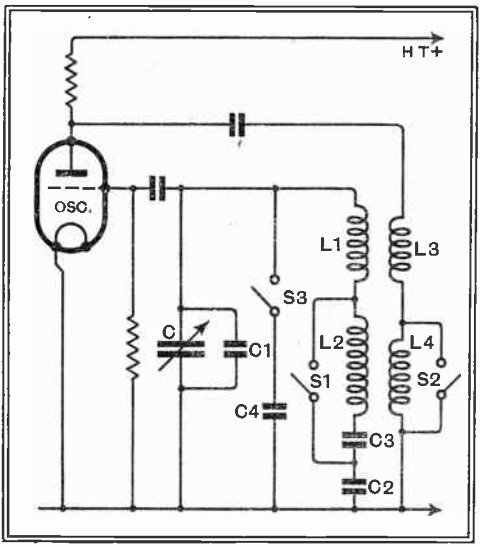
We have seen that the intermediate frequency is in all usual cases equal to the difference between the signal frequency and the oscillator frequency. With an I.F. of 450 kc/s the oscillator must therefore be tuned to a frequency either 450 kc/s greater or 450 kc/s less than the signal. If the oscillator frequency fo is higher than the signal frequency fs the intermediate frequency is $(fo - fs)$. If it is lower, the I.F. is $(fs - fo)$. At first sight it would seem a matter of indifference which of these alternatives were chosen. There are, however, marked practical advantages in making fo higher than fs.

Suppose the set is to tune from 1,500 to 550 kc/s (200 to 545 metres). Then, if of higher frequency, the oscillator must run from (1,500 + 450) to (550 + 450), i.e., from 1,950 to 1,000 kc/s. If, on the other hand, the oscillator is of lower frequency than the signal, it must run from (1,500 − 450) to (550 − 450), or 1,050 to 100 kc/s. The former range gives 1·95, the latter 10·5 as the ratio between highest and lowest frequency. Since even the signal-circuit range of 2·72 is often quite difficult to achieve, owing to the high minimum capacitances likely to be present in a finished set, the oscillator range from 1,950 to 1,000 kc/s would always be chosen in practice.

It is evident, since the frequency difference between signal and oscillator must be kept constant, that the oscillator must be tuned in a manner that is in some way different from the tuning of the signal-frequency circuits. There are three methods of tuning a superheterodyne. First, the signal-frequency circuits may all be made alike, and tuned by a multi-section gang condenser, leaving the oscillator to be tuned independently by another knob. The modern insistence on one-knob tuning is generally held to bar this method, though there is obviously no objection to it on purely electrical grounds.

At first sight it might appear impossible to tune the oscillator with a condenser-section identical with those tuning the signal-frequency circuits, because the required ratio of maximum to minimum capacitance is different. This difference, however, can readily be adjusted by putting a fixed capacitance either in parallel with the oscillator condenser to increase the minimum capacitance, or in series to

Fig. 139: Complete dual wave-range oscillator circuit, showing arrangement of series ("padding") condensers C2 and C3, and of parallel ("trimming") condensers C1 and C4. By correct choice of values for these, ganging may be made practically perfect



reduce the maximum. Having got the ratio of maximum to minimum correct in either of these ways, correct choice of inductance for the oscillator coil will ensure that it tunes to the correct frequency at the two ends of the tuning-scale.

In the middle, however, it will be widely out, but in opposite directions in the two cases. It is found that a judicious combination of the two methods, using a small parallel condenser to increase the minimum a little, and a large series condenser to decrease the maximum a little, will produce almost perfect " tracking " over the whole wave-band.

The resulting circuit is that of Fig. 139. Here C is a section of an ordinary gang condenser, and has at every dial reading the same capacitance as its companion sections tuning the signal-frequency circuits. With S_1 and S_2 closed, we have C_1 to increase the minimum capacitance and C_2 to decrease the maximum, their relative values being critical for accurate ganging.* Opening S_1 increases the inductance of the tuned circuit to enable the long-wave band (150 to 300 kc/s) to be covered by the set, at the same time decreasing the series condenser to the resultant of C_2 and C_3. At the same time S_2 is opened to throw in the extra reaction winding L_4, and S_3 is closed to add C_4 to the minimum capacitance in the circuit. The arrangement as a whole is shown, for simplicity, with a triode as oscillator, but it is equally suitable for use with a pentagrid or other specialized frequency-changer.

There is a third method of persuading the oscillator-circuit to tune at a constant frequency-difference from the signal-circuits. It consists simply in using a special multi-section condenser, in which the section tuning the oscillator has vanes shaped to give exactly the required results. Evidently this can be done only for one wave-band; for the long waves, therefore, the auxiliary fixed condensers must again be used. The circuit is that of Fig. 138 with C_1 and C_2 omitted, it being understood

* For formulæ to compute values and residual errors, see *The Wireless Engineer*, February, 1932, p. 70

that C is now no longer identical in capacitance with its fellows tuning the signal circuits, but has specially shaped vanes. The second method has now become practically universal.

159. Whistles

Owing to the characteristics of the frequency-changer, a superheterodyne is susceptible to certain types of interference from which an ordinary set is free.

The most noticeable effect of these is a whistle, which changes in pitch as the tuning control is rotated, rather like the results of using a "straight" receiver in an oscillating condition, but generally less severe. There are many possible causes, some of which are quite hard to trace.

The best-known is *second-channel* or *image* interference. In the preceding Section we have seen that it is customary for the oscillator to be adjusted to a frequency higher than that of the incoming signal. But if, while reception is being obtained in this way, another signal comes in on a frequency higher than that of the oscillator by an equal amount, an intermediate frequency is produced by it as well. If the frequency difference is not *exactly* the same, but differs by perhaps 1 kc/s, then two I.F. signals are produced, differing by 1 kc/s, and the second detector combines them to give a continuous note of 1 kc/s.

An example will make this clear ; and for simplicity the I.F. will be made a round number, 100 kc/s. Suppose the station desired works on a frequency of 950 kc/s. When tuned to it, the oscillator is 100 kc/s higher, 1,050 kc/s, and yields a difference signal of 100 kc/s, to which the I.F. amplifier responds. So far all is well. But if a signal of 1,149 kc/s is also able to reach the frequency-changer it will combine with the oscillation to produce a difference signal of $1{,}149-1{,}050=99$ kc/s. The I.F. amplifier is unable to reject this (Sec. 161), so both are amplified together by it and are presented to the detector, which produces a difference frequency, 1 kc/s, heard as a high-pitched whistle. Slightly altering the tuning control alters the pitch of the note, as shown in the table :

TABLE

Set Tuned to : kc/s.	Oscillator at : kc/s.	IF Carrier due to 950kc/s signal (wanted)	IF Carrier due to 1,149kc/s signal (interfering).	Difference (Pitch of Whistle).
		kc/s.	kc/s.	kc/s.
945	1,045	95	104	9
946	1,046	96	103	7
947	1,047	97	102	5
948	1,048	98	101	3
949	1,049	99	100	1
949½	1,049½	99½	99½	0
950	1,050	100	99	1
951	1,051	101	98	3
952	1,052	102	97	5
953	1,053	103	96	7
954	1,054	104	95	9

Since both stations give rise to carriers falling within the band to which the I.F. amplifier must be tuned to receive one of them, this part of the set can give no protection against interference of this sort. That is why it is necessary to have preselector tuning circuits. At first sight it might appear that their task is easy, for the interfering station is twice the intermediate frequency away from the wanted station, and even with an I.F. as low as 100 kc/s that is 200 kc/s. But it must be remembered that (1) the interfering carrier may be thousands of times stronger than the one desired ; (2) the product of two carriers is always considerably stronger than one carrier and its sidebands, and (3) 200 kc/s is only 1 per cent. off tune at 20 Mc/s. The demand for short waves, and cheap receivers (and therefore a minimum of variable tuned circuits), led to the I.F. in common use being raised from about 110 to about 460 kc/s, giving a separation of nearly 1 Mc/s from the image frequency. The problem has thus become less serious on medium waves, but is still very present on short waves, resulting in most stations being received by the cheaper sets at two settings of the tuning knob. Whistles, however, are not as bad as one might imagine, because the number of powerful stations within a given frequency band is less than on medium waves.

Another form of interference, much more serious if present, but fortunately easy to guard against, is that due to a station operating within the I.F. band itself. Clearly, if it is able to penetrate as far as the I.F. tuning circuits it is amplified by them and causes a whistle on *every* station received. Again, a good preselector looks after this ; but the 550 kc/s end of the medium waveband may be dangerously close to a 465 kc/s I.F. If so, a simple rejector circuit tuned to the I.F. and placed in series with the aerial does the trick.

The foregoing interferences are due to unwanted stations. But it is possible for the wanted station to interfere with itself ! When its carrier arrives at the detector it is, of course, always at intermediate frequency. The detector, being a distorter, inevitably gives rise to harmonics of this frequency ; that is to say, currents of twice, thrice, etc., times the frequency. If, therefore, these harmonics are picked up at the aerial end of the receiver, at the same time as it is tuned to a station working on nearly the same frequency, the two combine to produce a whistle. It is easy to locate such a defect by tuning the set to two or three times the I.F. The cure is to by-pass all supersonic frequencies that appear at the detector output, preventing them from straying into other parts of the wiring and thence to the preselector circuits (Sec. 100).

Oscillator harmonics are bound to be present, too ; and are a possible cause of whistles. Suppose the receiver is tuned to 220 kc/s and the I.F. is 465. Then the oscillator frequency is 665, and that of its second harmonic is 1,330 kc/s. If now a powerful local station were to be operating within a few kc/s of 865, its carrier would combine with the harmonic to give a product beating at an audio frequency with the 465 kc/s signal derived from the 200 kc/s carrier. Such interference is only likely to be perceptible when the receiver combines poor preselector selectivity and excessive oscillator harmonics.

The exclusion of most of the varieties of interference peculiar to the superhet is fairly easy so long as there are no overwhelmingly strong signals. But if one lives under the shadow of a transmitter it is liable to cause whistles in a number of ways. Besides those already mentioned, an unwanted carrier strong enough to force its way as far as the frequency-changer may usurp the function of the local oscillator and introduce signals of intermediate frequency by combining with other unwanted carriers. Any two stations transmitting on frequencies whose sum or difference is nearly equal to the I.F. may cause interference of this kind; and with the additive types of frequency-changer are actually known to do so. But a perfect multiplicative frequency-changer does not yield a combination frequency merely by mixing the two, and so is immune. The designer also guards against the danger by a suitable choice of I.F., provided that the frequencies of all pairs of stations that might be "local" are known. And, of course, preselector selectivity comes to the rescuc once more.

The foregoing list of possible causes of interference is by no means exhaustive; but it is only in exceptional situations that whistles are conspicuous in a receiver of good modern design. Some of the very cheap models, however, tend to revive a fault that used to be characteristic of the early superhets—the radiation of energy from the oscillator. One listener may interfere with another tuned to a station separated in frequency by the I.F. If he is on the long wave band, say 200 kc/s, his oscillator, on 665 kc/s will, if radiating strongly enough, interfere with a neighbour tuned in to, say, 668 kc/s on the medium band.

CHAPTER 16

TUNING CIRCUITS IN THE I.F. AMPLIFIER

160. The Task of the I.F. Amplifier

THE I.F. amplifier of a superhet. has to perform exactly the same duties as the R.F. amplifier of a " straight " set. It is really a fixed-tune R.F. amplifier which derives its signal not from the aerial direct, but from the frequency-changer, since this is the point at which the I.F. currents first appear. Just as in the case of the R.F. amplifier, the problems concerned consist mostly of the design of the tuned circuits involved.

The double advantages of fixed tuning and of having to deal with signals of comparatively low frequency completely transform the problem. The fact that our tuning is to be fixed allows us to use more tuned circuits without extra complication, and also to make careful adjustments that could never possibly hold constant over a waveband. The lower frequency, as we saw in Secs. 129 and 131, means that we shall have at our disposal coils of much higher L/*r* ratio—and hence of much higher selectivity—than we could possibly hope for when dealing with our signals at frequencies round about the 1,000 kc/s mark. We therefore set out, from the beginning, to attain a much higher standard of selectivity than we should dream of attempting in the design of a signal-frequency amplifier.

161. Characteristics of I.F. Coils

The most-used intermediate frequency is 450 kc/s, or values not far removed from this. Previously, 110 kc/s was usual. Experience shows that the values of L/*r* set in the table below can be achieved, even with comparatively small coils, with the various types of winding indicated. The figures make rough allowance for the damping effects of valves and other components connected, in the finished set, across the tuned circuits.

Frequency (kc/s)	Type of coil	L/r Henrys and Ohms
450	Solid wire, air core	25 to 30 $\times 10^{-6}$
450	Litz., air core	40 to 50 ,,
450	Litz., iron core	70 to 80 ,,
110	Solid wire, air core	50 to 60 ,,
110	Litz., air core	Up to 140 ,,

As mentioned in Sec. 129, these figures will rise or fall with the dimensions of the coil, so that they are necessarily only approximate. In addition, for a given frequency they depend on the value of L, growing less as this is increased owing to the fact that the series resistance r equivalent to dielectric loss or other forms of parallel damping is proportional to the *square* of the inductance.

It is not easy, unless one is very familiar indeed with the implications of these figures, to draw any immediate conclusions from them. We will therefore assume that we are called upon to design the I.F. coils for a superheterodyne that includes one stage of I.F. amplification. A typical circuit for the relevant part of the receiver is given in Fig. 140, where it will be seen that each of the two I.F. couplings includes two tuned circuits, making four in all. One at each point would suffice to provide the necessary coupling between valves, but as we know (Sec. 127) that the larger the number of tuned circuits the better the compromise between selectivity and high-note reproduction, this minimum is doubled.

To get an idea of the meaning of the L/r values just given, we will draw two overall resonance curves for four *cascaded* tuned circuits, one curve corresponding to circuits of $L/r = 140$,* and one to circuits of $L/r = 25$, these being

* *Micro*henrys and Ohms—which allows us to drop the cumbersome "$\times 10^{-6}$".

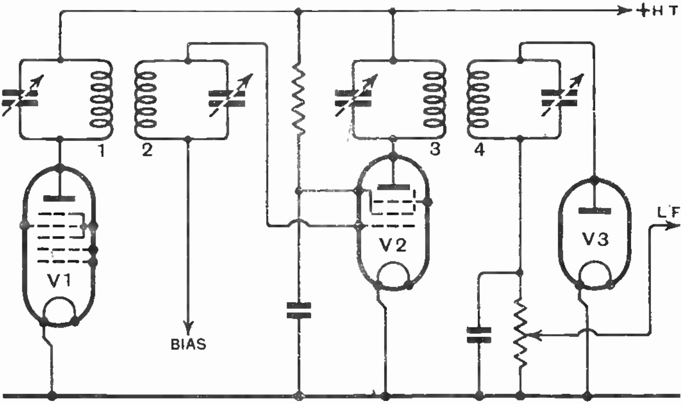


Fig. 140 : Skeleton diagram of single-stage I.F. amplifier. V_1 is the frequency-changer, V_2 the I.F. amplifying valve proper, and V_3 the detector

the highest and lowest figures in the Table. The curves, drawn from the data-curves of Fig. 115, are reproduced in Fig. 141.

The inner one, corresponding to $L/r = 140$, shows the most impressive selectivity—but also shows the most appalling loss of high notes. At 3 kc/s off tune (3,000 cycles audio) the response is little more than one-thousandth of that corresponding to the carrier (and the lowest notes).

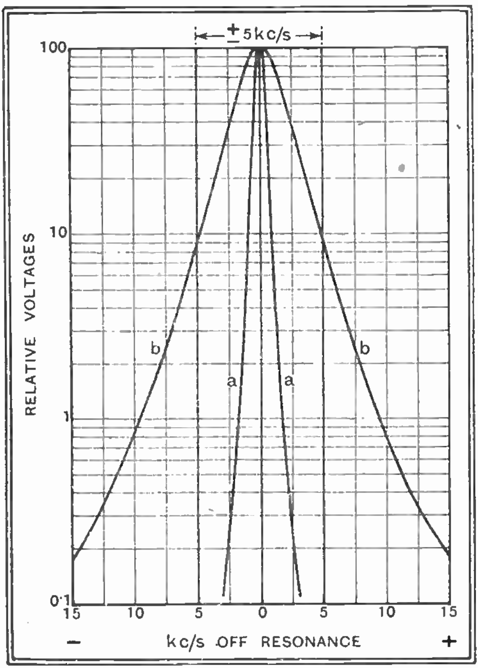
The outer curve, corresponding to $L/r = 25$, is more reasonable, being nearly 100 times down at 9 kc/s ; selectivity will be good, while at 5 kc/s (5,000 cycles audio) the response is still one-tenth of that for the bass. Even this curve, if realized in a receiver, would give very "boomy" and deep-toned reproduction of music, badly lacking in the life-giving high notes.

162. The Tuned Filter

The curves of Fig. 141 have been worked out on the assumption that the tuned circuits are in cascade, by which is meant that each retains its own individual resonance curve, unmodified by the presence of the others. But to pass energy through the intervalve couplings of Fig. 140, some coupling has to be provided by which this energy can pass from circuit 1 to circuit 2, and from 3 to 4. This is done by mutual inductance, the second coil lying in the magnetic field of the first. The process is analogous to that by which energy is transferred from an aerial to a tuned winding by coupling to the latter a few turns of wire connected between aerial and earth. But in the present case there is a difference—*both* circuits are tuned to the frequency of the currents supplied to them.

In such a case each circuit reacts upon the other, and each modifies the other's resonance curve. There emerges a new joint resonance curve, with characteristics that we have not yet discussed. This effect can equally be had by providing coupling of any other sort between the two tuned circuits. Fig. 142 shows three methods of coupling that are frequently used; in any one of the three cases the complete two-circuit system is known as a *filter*, or *band-pass filter*. More elaborate structures containing more than two tuned circuits, can be built up, but in

Fig. 141 : Overall resonance curves of four tuned circuits in cascade. *a* $L/r = 140$. *b* $L/r = 25$



ordinary wireless practice the use of tuned filters is generally restricted to a simple two-member combination such as that described.

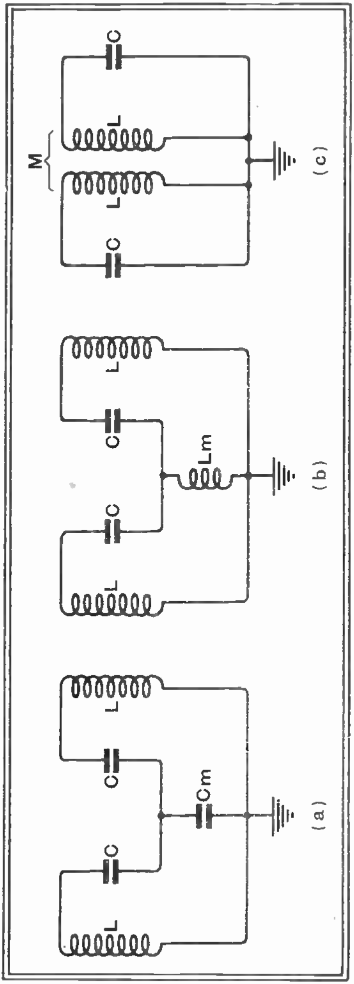


Fig. 142 : Three common types of tuned filter. Coupled in *a* by C*m*, common to both tuned circuits ; *b* by L*m* replacing C*m* ; *c* by mutual inductance M between the coils themselves

We have seen that the resonance curve of a single tuned circuit is determined entirely by the ratio L/*r*. In a filter we have a second variable in the coupling between the coils, which determines the degree of "spread" round the peak.

If we denote by X the reactance of the coupling element (C*m*, L*m*, or the mutual inductance M in Fig. 142), then the effect of the coupling in modifying the resonance curve from that proper to the same to circuits in cascade depends upon the ratio X/*r*. If, therefore, we know the sharpness of tuning of the individual circuits, as given by L/*r*, and also the effect of coupling, as given by X/*r*, we can plot the complete resonance curve of a filter. The formula necessary for this is given at the end of this Chapter.

To investigate the nature of the curve, we will take the very practical case of two tuned circuits (one intervalve coupling in Fig. 140), each of which has $L/r = 40$. If the coupling between them is very weak, so that the reaction of one circuit upon the other is negligible, we get, for the two circuits in cascade, the innermost resonance curve

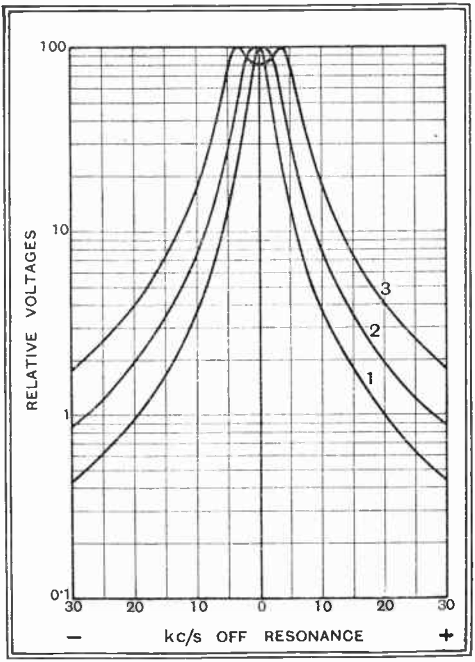


Fig. 143: Resonance curves of two tuned circuits, each $L/r = 40$. (1) Cascaded: $X/r = 0$. (2) Critically coupled $X/r = 1$. (3) Coupled to give overall band-width $\pm$ 5 kc/s. $X/r = 2{\cdot}04$. (See formula 6)

1 of Fig. 143, This shows a reduction of voltage to 19 per cent. at $\pm$ 4 kc/s from resonance, and to 4·5 per cent. at $\pm$ 9 kc/s. The weak coupling further ensures that even if a large voltage appears across the first coil, that across the second will be extremely small.

As the coupling between the two coils is increased by bringing them closer together, the voltage across the secondary increases and the peak of the resonance curve broadens, until at *critical coupling* the curve takes the shape shown at 2 in Fig. 143. The response at $\pm$ 4 kc/s has now risen to 44 per cent., thereby improving the transmission of high notes, but at the cost of a reduction in selectivity, the response at $\pm$ 9 kc/s now being 9·4 per cent. At this coupling the voltage across the secondary is half that which would appear across the primary used as simple tuned-anode coil.

With still closer coupling the voltage, at exact resonance, across the secondary begins to fall a little, while the joint resonance curve takes on the shape shown at 3. The rounded peak of curve 2 has now split up into two separate peaks, with a trough at the actual resonant frequency itself. The response at $\pm$ 4 kc/s is now 98 per cent. of the maximum, while at $\pm$ 5 kc/s it is equal to that at exact resonance. Selectivity has necessarily dropped further, the response at $\pm$ 9 kc/s having risen to 23 per cent.

It would appear that curve 3 offers a suggestion for a very satisfactory design. It provides a rising response up to 5 kc/s from resonance, thereby compensating for probable losses in other portions of the receiver, while at the same time giving selectivity which, by using a large enough number of pairs of circuits, might be made sufficiently high. In practice it is found that resonance curves of this type are very hard to realize, for differences in the L/*r* values of the two circuits generally lead to a curve in which one peak, being predominant, is brought exactly to resonance, while the other is represented by no more than a slight irregularity on one side or the other of a steeply falling curve. On the whole, it is safest for a designer to content himself with trying to get a peak only a little wider than that of curve 2, which represents the case of critical coupling and maximum gain.

163. Critical Coupling

Two circuits are critically coupled when the coupling is so close that the peak of the curve is just on the verge of breaking up into two separate peaks. This occurs when the coupling reactance X is made equal to the high-frequency resistance *r* of either of the circuits (assumed identical), or when the *relative coupling* X/*r* is made equal to 1. Naturally, the higher *r* is made the broader will be the peak, since raising *r* flattens the tuning of each individual circuit and at the same time involves an increase in X to maintain coupling at the critical point. A rapid

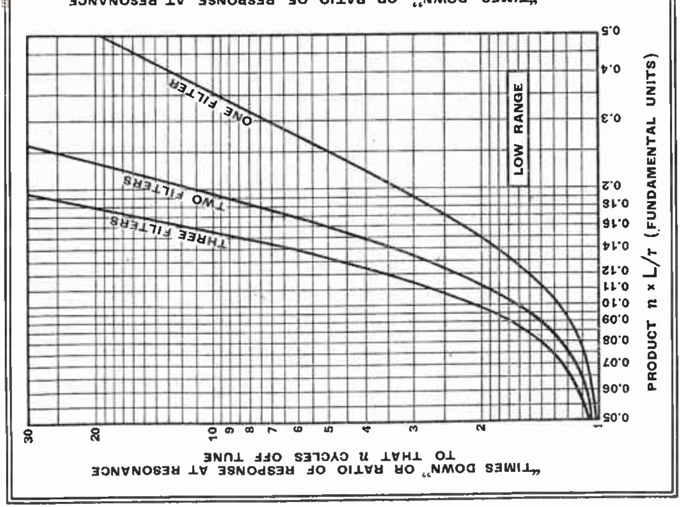


Fig. 144a: Design-curves from which overall resonance curves of one, two, or three *critically coupled* filters may be found if L/r for the individual circuits is known

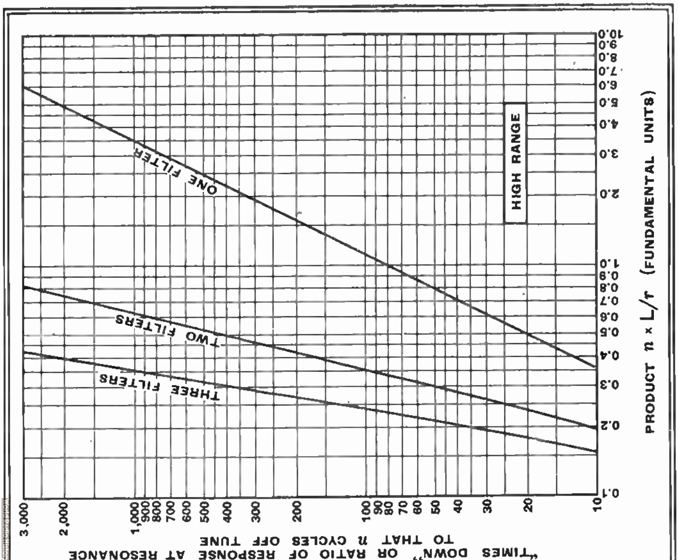


Fig. 144b: Continuing Fig. 144a to higher values

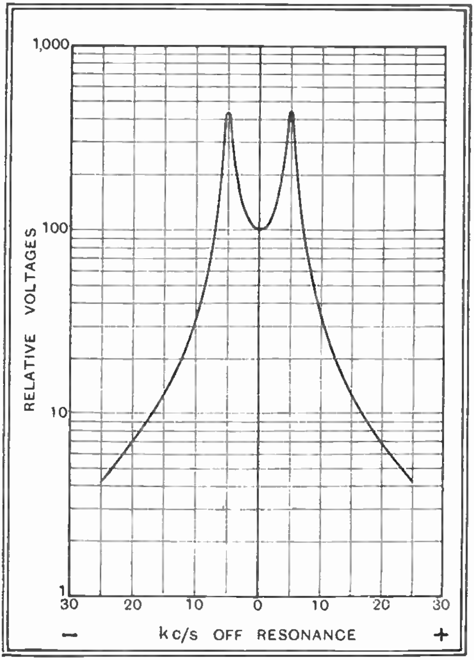


Fig. 145 : Showing how "rabbit's ears" develop when an attempt is made to broaden the peak by closely coupling coils of high L/*r*. L/*r* = 140. X/*r* = 79 (formula 3)

estimate of the width of the peak can be made by dividing L/*r* for the circuits concerned into 0·15, which gives the number of cycles off tune at which the response has fallen to half that at resonance.

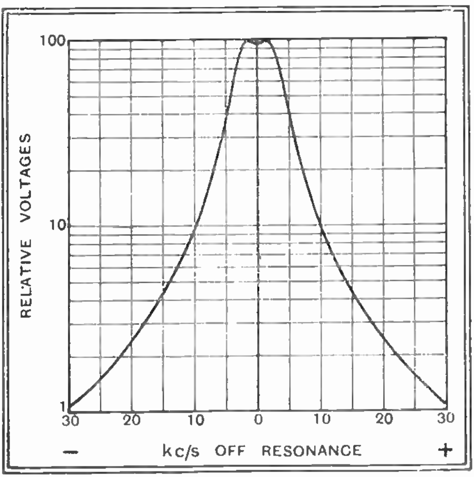
Thus for two circuits of L/*r* = 50, critically coupled, the curve would fall to half-height at $0{\cdot}15/(50 \times 10^{-6}) = 150{,}000/50$ cycles $= 3{\cdot}0$ kc/s off tune. Data for plotting rapidly a complete resonance curve for the particular case of critical coupling are given in Fig. 144. Here "times down" at *n* cycles off tune is plotted against the product $n \times$ L/*r*, the latter being in fundamental units (cycles, henrys and ohms). The curve applies to the simple case where the two tuned circuits are identical; in the case of any differcnce between them an approximation at least could be had by taking a mean value for L/*r*. This figure fulfils for a filter what the design curves in Fig. 115 do for circuits in cascade.

164. Coupling Closer than Critical

In the third (peaked) curve of Fig. 143 there are two peaks at about $3\frac{1}{2}$ kc/s either side of resonance. A curve of this type is just as easy to plot from the full formula as one for critical coupling, but short cuts are less simple. Owing to the difficulty of realizing such curves, we will do no more than refer the reader to formulæ, at the end of this Chapter, which give the number of cycles off tune at which the peaks occur, their height, and the number of cycles off tune at which the final fall of the curve outside the peak brings the response down again to equal that at resonance.

Attention is particularly drawn to the impossibility of combining a flat-topped curve with high selectivity by closely coupling a pair of very low-resistance circuits. Fig. 145 shows the curve of a filter in which each circuit has $L/r = 140$, coupled to give peaks at 5 kc/s off tune. Apart from the fact that the tuning of each circuit reacts upon that of the other to such an extent as to make the realization of the curve a matter of extreme difficulty, the great height of the peaks will lead the user of the finished superheterodyne so to tune his oscillator as to put the I.F. carrier, not in the trough, where signals will be quietest, but on one of the peaks, where the output of sound will, in the case shown, be twenty times as great.

Fig. 146 : Resonance curve of filter suggested as suitable for I.F. amplifier of Fig. 140. $L/r = 40$. $X/r = 1{\cdot}25$



165. Designing the Amplifier

We will suppose, therefore, that in supplying coils to the amplifier of Fig. 140 we shall content ourselves with a low L/r ratio and a relative coupling little tighter than critical. This will give us a curve that is not too selective for acceptable quality while keeping away from practical difficulties in tuning. Suitable values are $L/r = 40$, $X/r = 1 \cdot 2$ to $1 \cdot 3$, which give us (for one filter) the curve of Fig. 146. This is practically flat to $2\frac{1}{4}$ kc/s off tune, after which it drops away to a little less than half-height at $\pm$ 5 kc/s. At $\pm$ 9 kc/s it is nearly ten times down. Two such filters in cascade will give a resonance curve typical of that of the I.F. amplifier of the average modern superheterodyne.

The gain to be expected from the I.F. stage is very readily calculated. Since it depends on the dynamic resistance $(2\pi f L)^2/r$ (Sec. 59) of the tuned circuits, it can (theoretically) be raised to any desired value by choosing a sufficiently high value for L, of course keeping L/r constant at the chosen value. Let us suppose that the intermediate frequency is 450 kc/s, and that with an I.F. valve of slope 2·5 mA/v. we want a gain of 250 times from grid of I.F. valve to grid of detector. Since the coupling is close to the critical value at which the voltage output from the secondary is half that which would appear across the primary if only one coil were used, the gain with a one-coil coupling will have to be almost exactly double this figure, making 500 times. Dividing this by the slope of the valve gives the dynamic resistance required for the anode coil, which is therefore, 200,000 ohms. Knowing that $L/r = 40 \times 10^{-6}$, and $(2\pi f L)^2/r = 200{,}000$ ohms, we readily deduce* that L must be 625 μH, bearing in mind that $f = 450$ kc/s. This inductance we shall have to tune with 200 $\mu\mu$F, including strays.

If the pentagrid has a slope of 3 mA/v., the conversion conductance will be at best 1·5 mA/v. (Sec. 157), giving a gain of about $(200 \times 1 \cdot 5)/2 = 150$ times, reckoning from R.F. on modulator grid to I.F. on grid of I.F. valve. Since we have designed this stage to amplify 250 times, the overall gain from signal on grid of pentagrid to second detector will be 250×150, or about 37,500 times.

$$* \; L = \frac{R}{(2\pi f)^2 \, L/r}$$

166. Appendix : Filter Formulæ

The *resonance curve* of a filter is given by :

$$\left(\frac{Vo}{V}\right)^2 = \left(1 - \frac{158n^2p^2}{1 + q^2}\right)^2 + \left(\frac{25 \cdot 2np}{1 + q^2}\right)^2 \quad \ldots \quad \ldots \quad \ldots \quad (1)$$

where Vo = voltage at resonance
V = voltage at n cycles off tune
p = L/r (in henrys and ohms)
q = relative coupling X/r, or ratio of coupling reactance to coil resistance.

Critical Coupling occurs when $q = 1$ (See Fig. 142).

This gives maximum voltage on second coil, this voltage being half that which would have appeared on the first coil had it been the only one used. Still closer coupling reduces the mean voltage of the modulated carrier but little.

The resonance curve of a critically-coupled filter can be plotted from :

$$\left(\frac{Vo}{V}\right)^2 = 1 + 6245n^4p^4 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (2)$$

The data-curves of Fig. 144 are plotted from this, and provide a convenient short cut.

Peaked Curves. (q greater than 1.)

If peak is n cycles from resonance,

$$q^2 = 1 + 158p^2n^2 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (3)$$

and height of peak is given by :

$$\frac{V}{Vo} = \frac{1 + q^2}{2q} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (4)$$

(Use by finding q, by formula (3), from known L/r and desired n; then find V/Vo from formula (4).

Approximate short cut in a single stage : height of peak n cycles out from resonance is given by :

$$\frac{V}{Vo} = \frac{158n^2p^2 + 3}{4} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (5)$$

Overall Band-width

If it is desired that, at n cycles from resonance, the peak shall have been passed and the voltage shall have fallen again to the level of the trough at resonance, make :

$$q^2 = 1 + 79p^2n^2 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (6)$$

The rest of the curve can then be sketched by finding n for peak from (3) and height of peak from (4).

CHAPTER 17

AUTOMATIC CONTROLS

167. The Principle of A.V.C.

ALTHOUGH a few early superhets ended up with a grid detector and output stage, it is usual to take advantage of the high available pre-detector amplification to provide automatic volume control (A.V.C.) or, more correctly, automatic gain control. The purpose of this is, firstly, to cause all stations, strong or weak, to be reproduced at approximately equal volume, as set by the manual volume control knob ; and, secondly, to counteract the fluctuations in volume due to fading of the received signal. The principle of it is that the carrier reaching the second detector provides, by virtue of the process of rectification, a steady voltage which is used to bias back the earlier amplifying valves, so reducing their gain. For this reduction in gain to be effective it is evident that the peak voltage of the signal reaching the detector must be able to rise, without producing distortion, to a value equal to the bias required to reduce the gain of preceding valves to a low figure. This voltage may amount to 15 volts or more ; it is quite certain that no detector other than a diode can possibly handle voltages of this order.

168. Simple A.V.C.

Fig. 147 gives a simple A.V.C. circuit, in which the diode V_2 serves both as second detector and as generator of the A.V.C. voltages. The signal applied from the secondary of the I.F. transformer T across anode and cathode of V_2 is rectified in the usual way with the aid of the condenser C and the leak R, the latter being in the form of a potentiometer from which any desired portion of the total A.F. voltage across it can be conveyed to the A.F. amplifying valve. The flow of electrons through R on their way from anode to cathode of V_2 makes the " live " (unearthed) end of R negative to an extent substantially equal to the

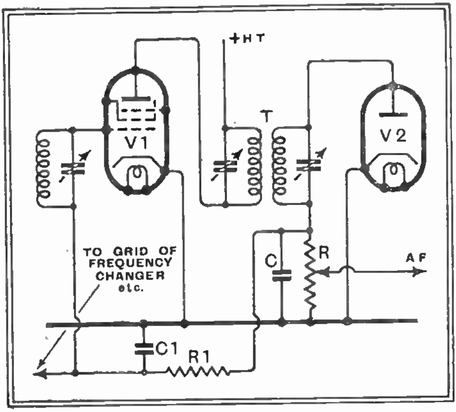


Fig. 147 : Skeleton circuit of simple A.V.C. arrangement. The D.C. voltage produced by signal rectification by V_2 is used for control bias

peak voltage of the applied I.F. signal that is driving the current. This voltage is fed back to the grid of V_1, the filter made up of R_1 and C_1 being interposed in the path to prevent carrier-frequency and audio-frequency voltages from being also fed back along the same path.

If we make the assumption that a bias of 15 volts on V_1 (and other pre-detector valves not shown in the diagram) will be required to reduce their amplification sufficiently to enable them to handle local-station signals, it is evident that when the station is tuned in, the peak I.F. voltage applied to V_2 must have this value. Further, it is evident that any station inducing a lesser voltage in the aerial will give rise to some lower voltage at V_2.

If the degree of A.F. amplification following V_2 is such that 5 volts (peak) of signal is required at that valve to provide full output at the loud-speaker, it will be impossible to obtain full-strength signals without at the same time applying 5 volts of bias to all pre-detector valves. This means that all stations weaker than this are prevented from giving full output, even though the set would have adequate sensitivity to receive them properly if it were not for the intervention of the A.V.C. system. If there are two controlled valves, and each has its slope reduced to one-tenth of its maximum value by the application of this bias, the sensitivity of the set will be one-hundredth of its maximum value.

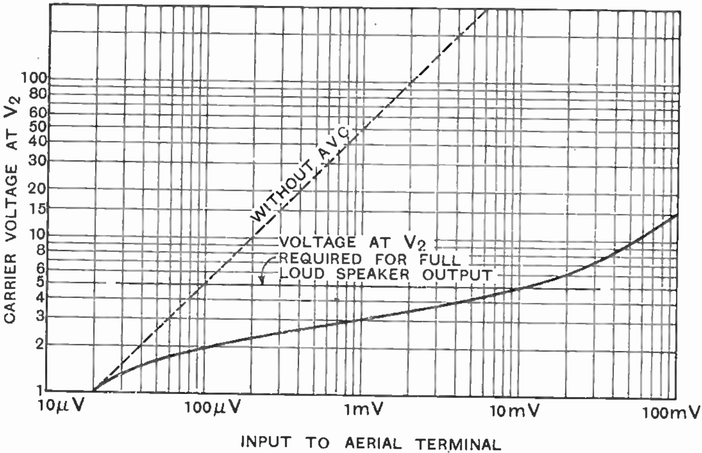


Fig. 148 : A.V.C. curve for system of Fig. 147. Note that if 5 v. at V_2 is wanted for full output, the A.V.C. is unnecessarily limiting output on all inputs from 20μV. to 10 mV

Fig. 148 shows, diagrammatically, the type of relationship between input signal and voltage at V_2 that would be given by a circuit like that of Fig. 147. As soon as the initial insensitivity of the detector is overcome, the rectified voltage applied as bias begins to reduce the sensitivity of the set, so that the climb in output with rising input becomes very slow. The dotted line shows how the output voltage would rise if, in the absence of the A.V.C. system, the amplification of the set remained constant irrespective of the signal applied.

It is fairly clear that the full useful sensitivity of the set could be regained if the A.F. amplification succeeding the detector were raised until 1 volt at V_2 provided signal enough to load up the output valve, for at this voltage the A.V.C. has barely begun to reduce the sensitivity. But if this were done, we should find that at the other end of the scale the output would rise to excessive values, for 15 volts bias, and with it 15 volts of signal, would still be produced by tuning in the local station. In spite of the A.V.C. system, drastic use of the volume-control would still be required on tuning from a near to a distant station, for the ratio of maximum to minimum power output would be 15^2, or 225 to 1.

169. Delayed A.V.C.

If we can arrange that the signal-voltage is always greater than the A.V.C. voltage, we can reduce this ratio very considerably. Suppose that the signal is allowed to rise to 5 volts before the A.V.C. system begins to operate; then, as 15 volts of bias will still be wanted for the local station, the signal it gives at the second detector must be 20 volts. On the assumption that the post-detector gain is so arranged that 5 volts at the detector fully loads the output valve, we now have a voltage ratio of 4 to 1 from loudest to faintest station within the range of A.V.C., or a power output ratio of 16 to 1, in place of the 225 to 1 of the circuit of Fig. 147.

This very considerable improvement can be realized in practice by the circuit of Fig. 149. So far as the signal-circuits are concerned, this is identical with Fig. 147. Detection now takes place at one anode D_1, of a double-diode valve, the leak being returned, as before, to cathode.

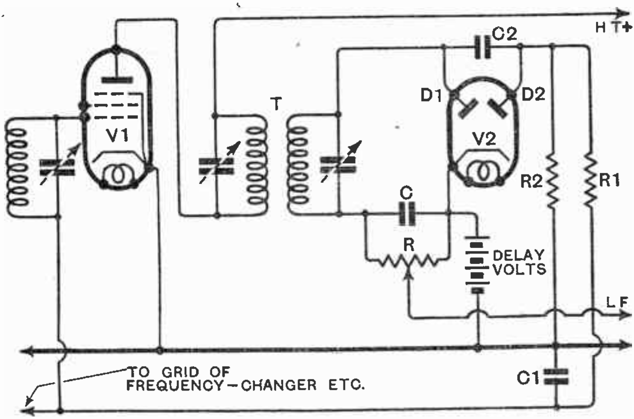


Fig. 149 : Modification of Fig. 147 to produce delayed A.V.C. Until the peak voltage of the signal exceeds the positive bias on the cathode of V_2, the A.V.C. system does not begin to operate

The signal is also applied, through the condenser C_2, to the second diode D_2, whose leak R_2 is returned to the earthline. By means of the battery shown, the cathode of V_2 is made positive with respect to earth, with the result that rectification at D_2 does not commence until the positive peaks of the I.F. signal run this electrode up to a voltage at least equal to that applied to the cathode.

If we make the cathode of V_2 positive by 5 volts and apply a 5-volt (peak) signal we can then adjust the post-detector gain until the rectified output just loads up the output valve. With this signal the A.V.C. diode D_2 is just about to begin to rectify; the signal is therefore allowed to build up to full output without interference from the A.V.C. system, which then immediately starts work and tends to prevent any further rise. For a set so adjusted, the A.V.C. curve, carried on to 15 volts bias (= 20 v. signal minus 5 v. delay) will be of the type shown in the lowest curve of Fig. 150.

The two other curves represent the response of sets having delays of 10 v., and 15 v. respectively, and it will be clear that as the delay increases so does the perfection

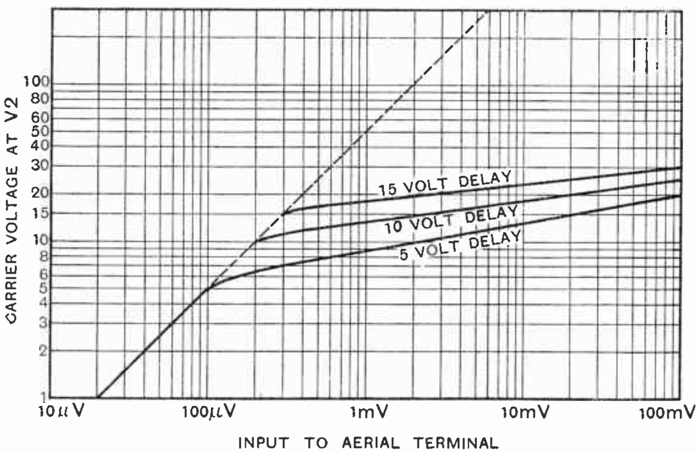


Fig. 150: A.V.C. curves for circuit of Fig. 149. Note that the larger the delay the flatter the curve, as explained in the text

of the A.V.C. system. With 15 v. delay the signal rises from 15 v. to 30 v.—a 2 to 1 ratio only—for the required increase in A.V.C. bias from zero to 15 volts. Higher delay evidently implies that we shall have to cut down the post-detector gain, so that the overall sensitivity of the set drops in proportion to the delay.

The simple A.V.C. system of Fig. 147 is practically never used, owing to the disadvantages described, but delayed* A.V.C. produced as in Fig. 149 is used in the majority of modern sets. In place of using the battery shown, the cathode of the double diode is made positive by connecting it to some point of suitable potential elsewhere in the circuit—usually to the cathode of the output valve.

Owing to the desirability of a large delay it is quite common to allow the signal-rectifier to supply the output valve direct, without intermediate amplification. If a high-slope indirectly-heated pentode is used, requiring about 4½v. peak signal, and a delay-voltage of 15 v. is provided, the output valve will be fully loaded on a carrier 30 per cent. modulated. Alternatively, the delay may be decreased a little, and enough amplification provided after the detector to allow a low-slope pentode or even a triode to be used as output valve. In this case it is usual to employ a double-diode-triode which, as its name implies, combines a double-diode for detection and A.V.C. with a triode for subsequent amplification, all being built into the same bulb.

170. A.V.C. Distortion

Either simple or amplified A.V.C. is liable to lead to distortion if the circuit, both of the A.V.C. system itself and of the I.F. amplifier, is not properly proportioned. It can be shown that if the audio-frequency load of a

* This is another radio term that might have been more wisely chosen. It must be understood that the "delay" is in voltage, not in time. There is actually a time delay in *all* A.V.C. systems, due to the components C_1 and R_1 in Fig. 147. In view of the possible rapid fluctuations of signal strength, on short waves especially, due to fading, this delay ought to be reduced to a minimum consistent with adequate elimination of audio frequencies. About a tenth of a second is suitable.

detector is less than the D.C. load, distortion occurs when the modulation depth, reckoned as a percentage, exceeds a hundred times the ratio of the two loads. In Fig. 147 the D.C. load of the detector is R, while the speech-frequency load is more nearly equal to R and R_1 in parallel (Sec. 94). If R is 0·25 MΩ and R_1 is 1 MΩ, which represent quite usual values, the audio-frequency load is 0·2 MΩ only, and distortion will occur if the modulation depth exceeds $\frac{0{\cdot}2}{0{\cdot}25} \times 100$, or 80 per cent.

A second source of distortion is found in the I.F. valve immediately preceding the detector, which in a set using simple A.V.C. is called upon to deliver a signal of the order of 10 to 15 volts when the local station is tuned in. With delayed A.V.C., the signal is even larger, being equal to the figure mentioned plus the delay voltage. To allow the last I.F. valve to pass on so large a signal it is not unusual to supply it with half only of the available A.V.C. voltage—which, on a strong signal, would bias the valve almost back to the bottom bend—but some risk of distortion still remains.

Both these sources of distortion can be avoided by using amplified A.V.C.

171. Amplified Delayed A.V.C.

When it is desired for any reason to work with a signal of the order of 1 volt at the detector, it is usual to provide amplified A.V.C., in which the rectified voltage is amplified before being fed back to earlier valves. This is done with the aid of a double-diode-triode in some such manner as shown in Fig. 151. As before, the signal is rectified by the diode D_1, with the leak R returned to cathode. The signal is passed for amplification to the grid of the triode, which is connected to the "live" end of R. The amplified signal is applied in the usual way to the grid of the output valve.

The cathode of the D.D.T. is connected, through a resistance R_3, to a point some 100 volts negative with respect to the general earth-line of the set. R_3 and R_4 are so chosen that with no bias on V_2 other than that generated

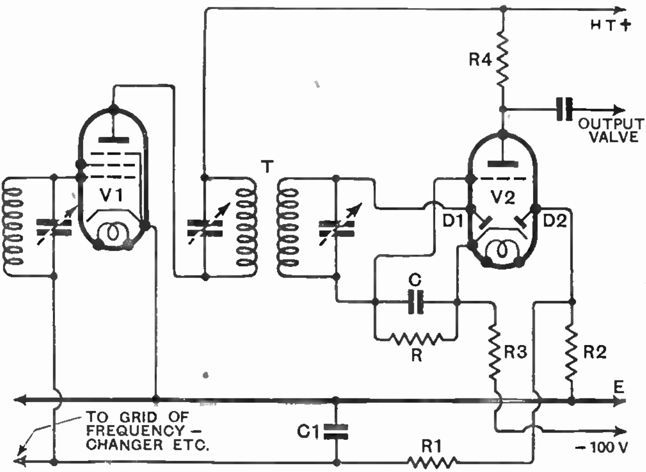


Fig. 151 : Amplified A.V.C., with delay. Initially positive, the cathode of V_2 is driven down to earth potential by the grid bias generated by rectification at D_1. At this point D_2 begins to draw current, and A.V.C. starts work

by grid-current through R the cathode is some 30 volts positive with respect to earth. When a signal is rectified by D_1 the resulting steady negative voltage, as well as the A.F. signal, reaches the grid of the triode. This negative bias reduces the anode current of the valve, thereby reducing the voltage-drop across R_3 and tending to make the cathode negative. If the amplification is thirty times, a one-volt signal on D_1 will drive the cathode from $+$ 30 v. to earth potential. A further half-volt will drive it down to $-$ 15 v. The diode D_2, connected to earth through the high resistance R_2, takes no current so long as the cathode is positive, but as soon as the cathode reaches earth-potential current begins, the resistance between cathode and D_2 drops to a negligible value, and D_2 follows the cathode downward in potential. Signals up to 1 v. on D_1 therefore generate no A.V.C. bias, and the set remains at full sensitivity, but by the time the signal reaches $1\frac{1}{2}$ v. the full bias of 15 v. is produced on D_2 and fed back to earlier grids in the usual way.

Thus, by this system, an even more level A.V.C. curve than that corresponding to a 15-volt delay in Fig. 150 can be produced from a 1-volt signal. Furthermore, the smallness of the signal ensures that the last I.F. valve shall never at any time be overloaded, while with the arrangements shown the audio-frequency and D.C. loads on the detector are identical, since the resistance R fulfils both functions. A.V.C. distortion is thus completely avoided.

By suitably increasing the positive potential of the cathode of the D.D.T., and increasing the signal voltage to correspond, almost perfect A.V.C. can be produced. It is possible to have a 10-volt delay (cathode at + 150 v., amplification 15) followed by a rise in A.V.C. volts to the required 15 on increase of the signal from 10 to 11 volts. The A.V.C. curve for a system of this sort is a very close approach to the ideal, in which the dotted line of Figs. 148 or 150 would be followed up to the point at which full loud-speaker strength was reached, after which there would be no further rise in output, no matter how greatly the input were increased.

172. Automatic Frequency Correction

To obtain good quality of reproduction from a superheterodyne receiver, particularly if the tuned circuits have high selectivity and the A.F. amplifier has a rising characteristic to compensate for loss of high notes, it is essential that tuning should be accurate. On the ordinary medium-wave band this implies care and a certain modicum of skill on the part of the user, while when receiving short waves not only is supremely exact tuning necessary, but it is also essential to have an oscillator that does not drift in frequency as the valve and other components warm up, or as a result of small fluctuations in mains voltage.

In either case the result of tuning slightly off the wavelength of the station being received is to over-accentuate the high notes carried by the sidebands of the incoming carrier, giving what is colloquially—and very descriptively—known as "side-band screech".

To avoid this many modern receivers are fitted with automatic frequency correction. This is operated from the output of the I.F. amplifier, and is so designed that whenever the I.F. carrier passing through the set departs from the frequency to which the amplifier is tuned, the control makes the necessary slight readjustment to the oscillator frequency that is required to bring the I.F. carrier back to its correct frequency.

The basis of the control consists of two sharply-tuned circuits arranged to peak one on either side of the nominal I.F. frequency and at a separation of about 4 kc. from it. These receive the signal from the last I.F. amplifier, and are connected to two separate rectifiers in such a way that the rectified currents are in opposition, as shown in Fig. 152.

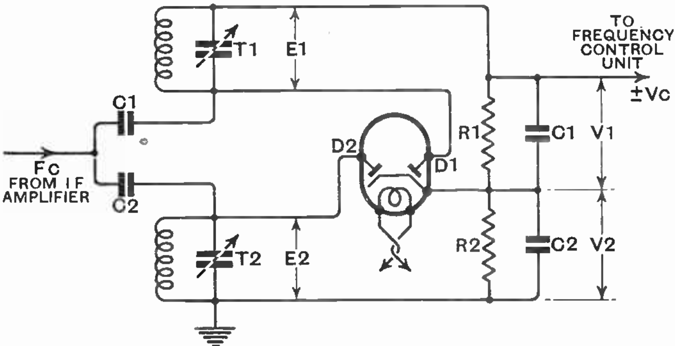


Fig. 152: The circuits T_1 and T_2, tuned one on either side of the correct intermediate frequency, pass to the double diode voltages which, when rectified, can be made to control oscillator-frequency

If the I.F. carrier F_c has the exact correct frequency, and the two tuned circuits peak at frequencies equally spaced on either side of it, E_1 and E_2 will be equal, and the rectified voltages V_1 and V_2 will also be equal. Being opposite in direction, the control-voltage developed will be zero. If F_c now approaches the frequency of the upper tuned circuit, E_1 will become greater than E_2, and V_1 will therefore exceed V_2, producing a resultant control voltage V_c that is negative in sign. Similarly, V_c will be positive if F_c drifts in the other direction, for E_2 will now exceed E_1, so that V_2 will be greater than V_1. This system, called the *discriminator*, thus provides us with a voltage that depends for its sign on the direction in which the I.F. carrier departs from its correct value, becomes greater when mistuning is increased, and falls to zero when tuning is accurate. A somewhat different arrangement, known as the phase discriminator, is now more commonly used.

This voltage can be used to control oscillator frequency in any one of several ways, of which perhaps the simplest consists of connecting grid and cathode of a "control valve" across the oscillator tuned circuit. If this valve has its gain controlled by the output from the circuit of Fig. 152, its input capacitance can be made to change sufficiently to provide the necessary small alteration in oscillator-tuning.

In setting up a circuit of this kind care has to be taken that the control is in the right direction, so as to increase oscillator frequency when it is too low and *vice versa*. Incorrect connection, remedied by interchanging earthed and live sides of the output of Fig. 152, results in the slightest mistuning being automatically increased so that as soon as a station is found the control tunes it out again.

Correctly connected, the control takes charge of the tuning as soon as the dial is so set as to bring in a carrier that will pass through the I.F. amplifier, automatically adjusting the tuning to the correct point for that particular station.

By means of a resistance-condenser combination, the action of the control is made to lag behind the reception of the carrier by several seconds in order that, in searching, the response of the set to manual tuning may be normal.

Sideband screech is combated in simpler sets, not provided with A.T.C., by driving the A.V.C. diode from the *primary* of the final I.F. transformer, at which point the selectivity is lower and consequently the A.V.C. bias produced at frequencies slightly off-tune (sidebands) is greater than if the secondary were used.

173. Automatic Selectivity Control

It is possible, by means of rather complex circuits, to devise schemes whereby the resonance curve of an I.F.

amplifier can be broadened or narrowed by automatic means. In general a strong signal, received from a nearby station, is heard without much interference from other transmitters, while when receiving a weak station other transmitters on neighbouring wavelengths are liable to interfere. The A.V.C. system, therefore, may be used to control selectivity, broadening the tuning curve, initially of high selectivity, on receipt of a strong signal. By this means a rough-and-ready automatic adjustment of the selectivity-quality compromise to suit changing conditions may be made, with the limitation that for all strong stations, whether interference is present or not, high quality and low selectivity is provided, while for all weak stations, even if no interference is present, the opposite adjustment is made.

A still more complex, but at the same time more satisfactory solution to the problem may be made on the lines of Fig. 152, with the difference that the auxiliary circuits are now tuned to the channels on either side of the required station—i.e., to frequencies 9 kc. higher and lower than the intermediate frequency. By reversing the connections of one detector so that the rectified voltages add, it becomes possible to narrow the selectivity curve of the receiver, initially made broad, whenever a signal is present on either of the channels adjacent to that being received. A disadvantage of this scheme is that interference is in any case weaker than the desired signal, so that it becomes necessary to provide two extra I.F. amplifiers, tuned 9 kc. on either side of the one dealing with the signal, to amplify the interference sufficiently to enable it to provide an adequate control-voltage from the interference-detecting system.

CHAPTER 18

TAKING POWER FROM THE MAINS

174. Heating Battery and Mains Valves

IN a battery-driven set the filaments of all valves are connected together in parallel, and the necessary power to heat all of them is derived from a single 2-volt accumulator cell. The filament current taken by the valves depends on the anode current they are likely to be called upon to deliver ; 0·1 amp. is usual for detector valves, screened valves for R.F. amplification may take 0·1 to 0·2 amp., and output valves usually 0·2 amp. at least. The power used for heating the filament of a valve is therefore from 0·2 to 0·4 watt, or a little more in some cases. An average accumulator will supply an ampere for some 20 hours on one charge (a " 20 ampere-hour " cell) ; this is equivalent to running a 3- or 4-valve set for some 40 hours, which may represent a week or a fortnight of ordinary use.

Valves designed for mains operation are of two types ; those intended for A.C.-driven sets and those meant for the " universal " sets that run from either A.C. or D.C. In the former class the heater usually consumes 1 amp. at 4 volts, though a 2-amp heater is quite usual for output valves. The power used for heating is thus 4 to 8 watts, or twenty times as much as is used in battery valves. These 4-volt A.C. valves are used with their heaters connected in parallel, the power for all the valves in a set being taken from a transformer which steps the voltage of the mains down to the required figure.

Allowing for loss in the transformer, the heaters of a 3-valve set (16 watts) could be energized for fifty hours for the cost of one "unit" (kilowatt-hour) of electricity, so that the currents taken, though large by battery-set standards, are not by any means uneconomic.

Where D.C. mains are used, or where it is desired to dispense with the transformer, the heaters of all the valves are connected in series across the mains. For the sake of economy the valves are designed to operate at a low current (usually 0·2 amp.), and the voltage across each at this current varies from 13 to 40 volts, according to the wattage it is deemed necessary to dissipate in the heater. The larger voltages, of course, are required by the valves taking the largest anode current, i.e., the output valves. A resistance of the right value to drop, at 0·2 amp., the voltage by which the mains exceed that required by the valves is included in the circuit as at R in Fig. 153 *b*.

In this arrangement the power in watts consumed by the filament circuit as a whole is equal to one-fifth of the voltage of the mains, irrespective of the number of valves. With more valves R is reduced, so that less power is dissipated in it and more in the valves.

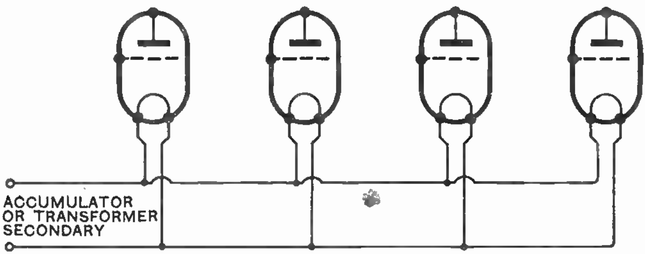
The greatly superior area of a cathode as compared with a filament, together with the fact that the whole of it is at the same potential, enables the mains valve to have a slope about double that of a corresponding battery valve. Further, the greater rigidity of a cathode allows the grid to be brought closer to it, this contributing further to high slope. One may, in consequence, quite fairly expect a mains set to be considerably more sensitive than a battery set of corresponding design.

175. Grid Bias in Mains Sets

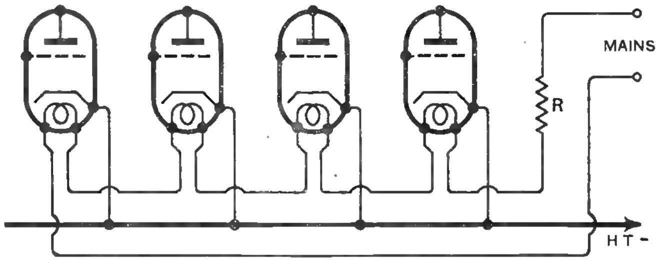
In the case of a battery set it is usual to provide a separate battery for providing the voltages at which the grids of the various valves are set. The positive side of this battery is connected to the negative side of the filament battery (LT –), and the grid return leads of the various valves are connected to suitably-chosen tappings on the battery, as shown in Fig. 132.

Bias in a mains set is derived in all cases from the H.T. supply. If we insert a resistance (R, Fig. 154 *a*) between the cathode of a valve and H.T. negative, the whole space-current I of the valve* has to flow through it. In so doing it makes the cathode positive by IR volts with respect to earth. If now we return the grid to earth, as in the diagram, it will be negative to the extent of IR volts with respect to the cathode.

The condenser C is placed across R because the latter is included both in the anode-cathode and in the grid-cathode circuits of the valve. Amplified signal currents in the anode circuit, in flowing through it, will therefore



(a)



(b)

Fig. 153 : Diagram *a* shows method of heating filaments of battery valves or heaters of A.C. valves in a 4-valve set. All valves require the same voltage. In diagram *b*, which shows D.C. or universal valves with heaters in series, all take the same current. Note that in spite of different potentials of heaters all cathodes can be joined to H.T.—

* The " space current " is the total of all currents to anode, screen, suppressor, and any other electrodes there may be.

introduce a signal-voltage back into the grid circuit. This voltage is in opposition to that due to the original signal; "degeneration", or reduction of amplification by reverse reaction, therefore, occurs. By making C large enough (50 μF is common) this effect can be entirely avoided except for the very lowest audio-frequencies.

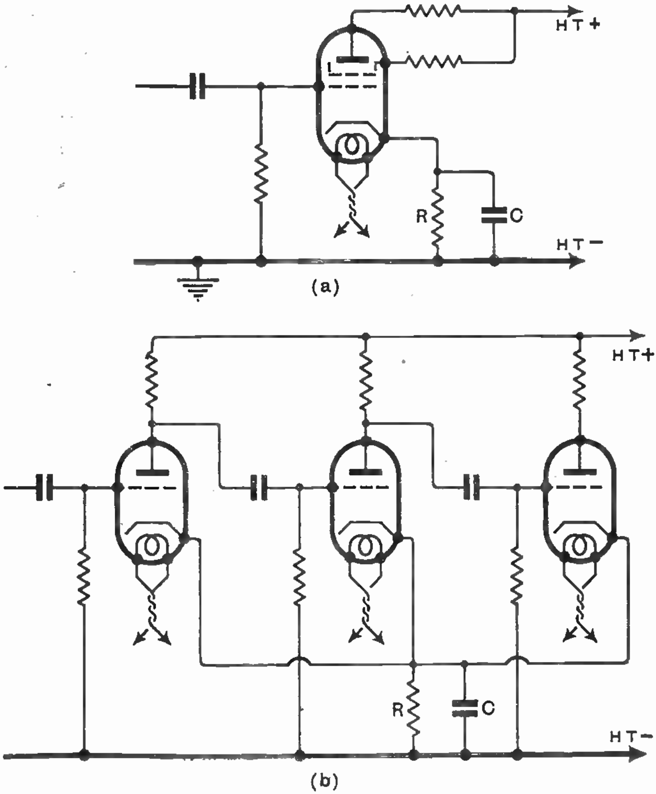


Fig. 154: Diagram *a* shows a true self-bias circuit, where the passage of anode current through R makes the cathode positive with respect to H.T.— ; in consequence, the grid is made negative with respect to cathode. In diagram *b* all valves are similarly biased to an equal extent by the voltage drop across R in their common cathode lead

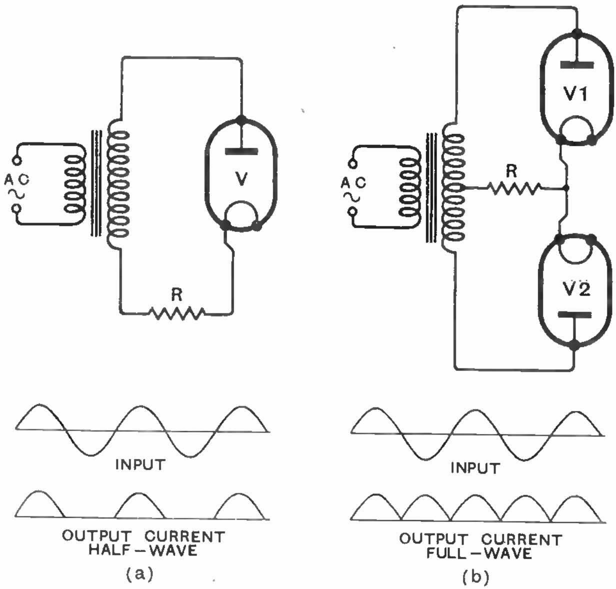


Fig. 155 : Half-wave and full-wave rectification. In *a* V acts literally as a valve, suppressing alternate half-waves ; in *b* there are in effect two transformer secondaries, phased so that the pulses through V_1 come between those through V_2

When several valves in a set require the same bias, some saving of components results by connecting all their cathodes together and inserting R and C in the common cathode circuit, as in Fig. 154 *b*. (In this circuit anode and grid resistances stand for couplings in general). Alternatively, R may be placed in the common negative lead of the set ; this is useful where the valves to be biased are controlled by the A.V.C. system, for the change in their space current is a small proportion only of the total current of the set.

176. Anode Current from the Mains

Supplied by an ordinary dry " H.T. battery ", one may reckon a unit (kilowatt-hour) of electricity to cost some

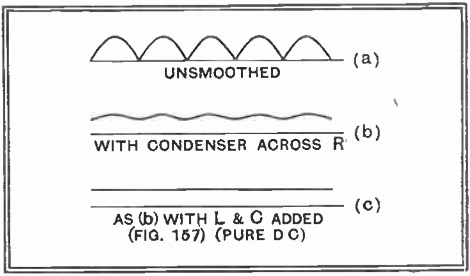


Fig. 156 : Showing how the alternating component can be removed from the crude rectified current

thirty shillings at least. From the mains, even allowing for all losses in conversion, one shilling would be a generous estimate. One can therefore afford in a mains-driven set to use plenty of anode current, which means, in turn, a more generous output and less need to run the last valve permanently on the verge of distortion due to overloading.

The power is there ; the problem lies in making use of it.

177. Rectification and Smoothing

The fifty-cycle alternations of A.C. mains, if allowed to reach the signal circuits of the set by any path, will produce a 50-cycle note (deep hum) in the speaker. Before we can use it we therefore have to convert the current *completely* from alternating current to direct.

This conversion is known as *rectification*, and is performed with a two-electrode valve. Fig. 155 shows, better than could any amount of description, how *half wave* rectification (at *a*) and *full-wave* rectification (with two valves, as at *b*) are carried out. In either case the result is a series of pulses of current all in the same direction, which we can equally well describe as a direct current with an alternating current superposed upon it. Freedom from hum can only be had if the alternating component is completely suppressed.

If we place a condenser of large capacitance across the resistance R a good deal of the alternating current will be diverted through the condenser. As a result the current through R is *smoothed*, taking on a wave-form such as that in Fig. 156 *b*. This, it is evident, is a much nearer approach to pure direct current, which would be represented by a horizontal straight line. By adding a choke and a second condenser to the circuit, as shown in Fig. 157, the small residue of alternating current is almost entirely removed, and the system of that figure can very satisfactorily be used to supply anode current to a set.

It is to be noticed that the full-wave rectifier V, containing a cathode and two separate anodes, draws its filament or heater current from the same transformer that provides the anode current. For the heaters of the various valves in the set proper still another winding would be used, a common primary winding energizing, through the iron core, as many secondaries as may be required for the entire receiver.

In a battery set suitable voltages for the screens of S.G. valves, and for any other points requiring less than the maximum voltage, can be obtained by connecting to suitable tapping-points on the battery. Since there is only one voltage available in a supply unit such as that drawn, it becomes necessary to utilize the voltage-drop across a resistor if lower voltages are required. For screen-grid valves it is usual to provide a potentiometer consisting of two resistances connected in series across the whole voltage, and to connect the screen, together with its by-pass condenser, to the junction point of the two. For screened pentodes, in which the screen current is larger and varies

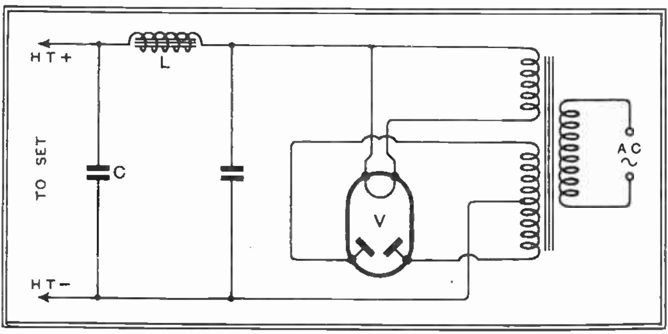


Fig. 157: Complete H.T. supply system for A.C. set. (Note that the set itself replaces the load resistance R of Fig. 155)

less from valve to valve, it is usually satisfactory to connect the screen through a resistance to the main positive line.

In the majority of mains-driven sets the loudspeaker is of the "energized" moving-coil type, requiring the dissipation of some five to ten watts in the windings of the electro-magnet used to provide the magnetic field in which the coil moves. The inductance of a winding of this sort is quite high, and it is convenient to place it in series with the main H.T. lead in such a way that the total anode current drawn by the set passes through the winding and energizes it. It then serves also as a very satisfactory smoothing choke, taking the place of that shown in Fig. 157. The voltage dropped across it is made up by increasing the alternating voltage applied to the rectifier V.

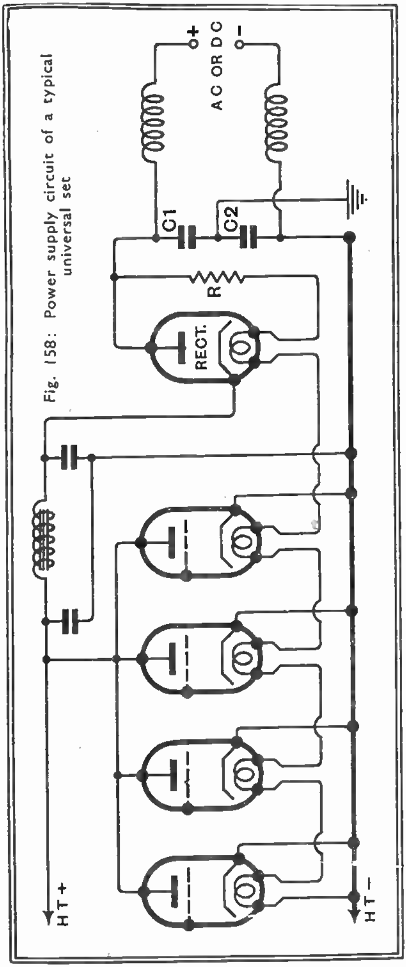


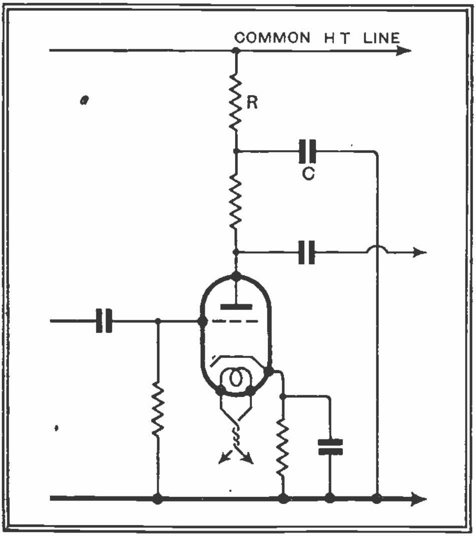
Fig. 158: Power supply circuit of a typical universal set

178. D.C. and Universal Sets

In the case of receivers intended to be run on D.C. mains, rectification is no longer necessary, but owing to the fact that the current is generated by rotating machinery it contains a small alternating component. To prevent hum this must be removed ; the smoothing choke and condensers are therefore retained. In this type of set it is usual to put the speaker field directly across the mains, since too much of the available voltage is wasted if it is used as a smoothing choke. Sometimes, however, it is used in place of R (Fig. 153 *b*) in series with the heaters.

Universal sets, running from A.C. or D.C. mains, are arranged as in Fig. 158. As in D.C. sets, the heaters of the receiving valves are in series ; in addition, there is included in the circuit the heater of an indirectly heated rectifier. On A.C. mains this acts as a half-wave rectifier, while on D.C. mains it is a "passenger", doing no more than add a small resistance in the H.T. line. Both universal and D.C. sets are inclined to be a little limited in output on account of the comparatively low anode voltages available ; in neither case can a transformer be used to raise the voltage above that of the mains.

Fig. 159: Decoupling a valve from the H.T. line is performed by inserting R to block signal currents, and providing C to give them a path back to earth



The R.F. chokes and small condensers C_1 and C_2 included in Fig. 158 are very necessary in both universal and D.C. sets ; they prevent high-frequency disturbances due to electrical apparatus connected to the mains from reaching the set. In an A.C. receiver their place is usually taken by an earthed screen between primary and secondary of the transformer.

179. Decoupling

The impedance to signal-frequency currents of the smoothing and rectifying circuits in a mains-driven receiver is considerably higher than that of a battery in good condition. Since this impedance is common to the anode circuits of all valves in the set it tends to couple them all together, and may set up instability of one sort or another. When this unfortunate state of affairs arises, *decoupling* is resorted to. As shown in Fig. 159, a resistance R is inserted in the anode circuit of such valves as require it, and a condenser C is connected from the high-potential side (from the signal-frequency point of view) of this resistance to earth. Condenser C then completes the anode circuit for signal-frequency currents, while R prevents any appreciable portion of these currents from finding their way back into the anode-current supply system. The larger C and R, the more complete the decoupling, which depends on the product CR.

CHAPTER 19

RADIATION AND AERIALS

180. Bridging Space—Radiation

IN the first chapter of this book the processes of radio broadcasting were very briefly traced from start to finish; but since then we have concentrated on the link in the chain nearest the listener—the receiver, beginning at the point where it is connected to the aerial. The transmitter, although the details of its design are outside the scope of the book, is based on the principles outlined in the first few chapters. But the stages between —transmitter-aerial, space, receiver-aerial—involve ideas that so far have only been hinted at.

It has been implied that even when sound waves have been converted into electrical waves these electrical waves are not of themselves able to travel large distances in empty space, although they can do so along wires. And electric current, as we have seen (Sec. 9), consists of electrons in motion, and these electrons are available in immense numbers in metals but are comparatively absent in insulators such as air. It is difficult to force them to cross even a fraction of an inch of air, so obviously hopeless to expect them to go hundreds or thousands of miles.

But we have seen (Sec. 8) that the space surrounding any electric *charge* is pervaded by a mysterious influence called an *electric field*, which is able to attract or repel other electric charges separated by air, or even by empty space. Also (Sec. 19) an electric *current* causes another mysterious influence, a *magnetic field*, capable of inducing currents in circuits some little distance away.

A *fixed* electric charge, or charges in *steady* motion (an electric current), set up forces in the space around, but these forces diminish very rapidly as one moves away, and at a short distance are too small to detect. This is analogous to a perfectly steady current of air, which may cause a slight local increase in air pressure, but is noiseless. However, when even a slight pressure of air is interrupted, say, 1,000 times per second, it is audible perhaps miles away. It sets up air *waves*. In the same way, if an electric current is rapidly *varied* in strength it sets up waves that radiate outwards through space with the speed of light and die away much less with distance than the induction fields.

An electric current cannot be made to *increase* rapidly for ever, so to keep it varying continuously at a rapid rate it has to be alternately made to rise and fall, as in Fig. 27. The current supplied for lighting, heating and power does this, it is true ; but only 50 times per second, and that is too slow to cause appreciable radiation. It corresponds to waving a fan slowly from side to side in the air—production of sound waves is negligible. But attach the fan to something that makes it vibrate 1,000 or so times per second and it will be painfully audible. A child's vocal organs can utter piercing sounds because they are high-pitched—many vibrations a second—and are much more audible than even a powerful singer's voice in the extreme bass.

The electric currents resulting from the action of sounds on a microphone are, of course, of the same frequencies as the sounds ; and even when as high as thousands per second they are too slow to stir up waves capable of travelling a useful distance. Hence the elaborate processes of modulation and detection, to make use of a carrier wave having a much higher frequency than the sounds. Although some radio communication is carried on by means of carrier waves alternating a few tens of thousands of cycles per second, radiation is much more complete when the frequency runs into millions.

181. Electromagnetic Waves

The waves consist of a strange and complete union of electric and magnetic fields, so complete that each depends entirely on the other and would disappear if it were destroyed ; and are therefore termed electromagnetic waves. They break away from the circuit that gave them birth, and keep on travelling even if the current in the circuit stops.

If one end of a rope is waggled rapidly up and down, a *vertical* wave travels *along* the rope ; that is to say, at right angles to the direction of waggling. This analogy represents the electric half of an electromagnetic wave. To complete the picture the magnetic part has to be imagined at right angles to both, and therefore in a side-to-side direction. Fig. 160 shows part of a wire in which an electric current must be supposed to be alternating, up and

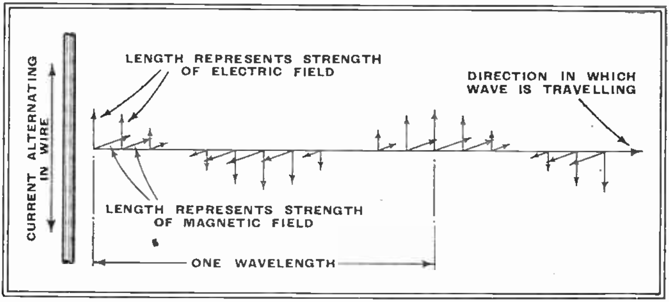


Fig. 160 : Representing one and a half cycles of radiated electro-magnetic wave moving outwards from a wire carrying oscillatory current

down. Then the electric part of the wave is parallel to the wire and alternates as indicated by the varying lengths of the vertical arrows : the magnetic part is at right angles and is indicated by horizontal arrows (actually drawn as if viewed in perspective) ; and the combined wave moves outward from the wire and therefore its direction of motion is at right angles to both the fields. It is extraordinarily difficult to depict this three-dimensional phenomenon on paper, for the wave front actually forms a complete cylinder expanding around the wire.

The electric field is measured in volts per metre, if close to the radiating wire, or microvolts per metre if far off. This voltage actually exists in space, but naturally can cause a current to flow only if it impinges on a conductor, such as a wire. The voltage may be considered to be due to the cutting of the wire by the magnetic field (Sec. 23), just as the magnetic field may be considered to be due to the moving electric field (Sec. 18). But a *horizontal* wire will have no voltage in the direction of its length when struck by waves due to a vertical wire radiator; and vice versa. This distinction is known as *polarisation*; the waves due to a vertical wire radiator are conventionally called vertically polarised, that being the direction of their electric field. If a vertical wire half a metre long is held in vertically polarised radiation of strength 50 microvolts per metre, then an E.M.F. of 25 microvolts will be induced in it. But if either the wire or the radiation is horizontal, no E.M.F. is induced. At intermediate angles to the horizontal, the induced E.M.F. is proportional to the sine of the angle. To induce the maximum E.M.F. in a horizontal wire it is necessary for the radiation to be horizontally polarised, which can be arranged by placing the radiating wire horizontally. If radiation starts off with a certain polarisation it does not follow that it will arrive with the same. If it has travelled far, and especially if on its way it has been reflected from earth or sky, it is almost certain to have become slightly disarranged, and at least part of it may be picked up by a wire at almost any angle.

182. The Coil Radiator

It is time now to examine the radiating circuit a little more closely. Up to the present we have considered only a short section of wire in which a radio-frequency current is flowing, and have conveniently disregarded the rest of the circuit which is necessary for its existence. In Sec. 30 it was shown how oscillating currents of almost any frequency can be set up in a very simple circuit consisting of a coil and condenser; and, later on, how by suitably connecting a valve to it the oscillations can be kept going continuously. We can fulfil the necessary condition for radiating waves, namely, causing a current to vary extremely rapidly—millions of times a second if need be. Yet such a circuit does not turn out to be a very efficient radiator. What is the trouble?

The difficulty is that for every short section of circuit tending to radiate electromagnetic waves there is another carrying the current in the opposite direction and therefore tending to cancel it out. Each turn of wire in the coil has to go *up* one side and *down* the other in order to complete itself; and, as the same current flows through the whole turn, the wave-producing efforts of the two sides of the turn are in opposition and largely counteract one another. Largely, but not altogether. Look at Fig. 161 and imagine it to be one turn of wire, for convenience square. Then you, looking at it, are an equal distance from equal lengths of wire carrying equal current and therefore radiating with equal strength; but as the current is always in opposite directions you never receive any net radiation at all. The same argument applies to the top and bottom sections. But now consider it from the point of view P. The down section is nearer than the up, and may therefore rather more than counteract the latter. Actually this disparity is quite negligible if P is miles away. But there is another thing to take into account. The radiation from the up section takes time—very little, it is true—to reach the down section on its way to P, and if the oscillations are very rapid indeed the current in the down section may have started to go up; in which case its radiation assists that coming from the previous up wire, and P will get the combined result. Even if the oscillation frequency is not high enough for a complete reversal, it may at least cause sufficient difference between the radiation from the two sections of wire to give a net result at P.

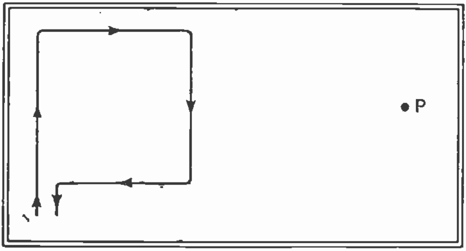


Fig. 161: From the point of view of the reader, the radiation due to different parts of a turn of wire carrying oscillatory current cancels out. Towards the point P, however, the different distances of the vertical sides of the square result in net radiation

The best result is obtained when the diameter of the coil is half a wavelength (or any odd number of half wavelengths), because then the radiative efforts of both sections pull together in the direction P—and also in the opposite direction. If the coil is 2 inches in diameter—about 5 centimetres—the best wavelength is 10 (or $\frac{10}{3}$, $\frac{10}{5}$, etc.) centimetres, the frequency being 3,000 Mc/s ! There are technical difficulties in producing such high frequencies, and even when they are produced the range is extremely limited. For most purposes, then, there are other reasons why the wavelength should be much longer. That being so, the obvious answer is to make the coil larger. As one sometimes wants to radiate waves thousands of metres long, this suggestion has its difficulties too, even bearing in mind that it is not essential for the diameter of the coil to be as much as half a wavelength in order to radiate a useful amount. For these and other reasons radiators in the form of a coil are comparatively seldom used.

How about the other element in an oscillatory circuit—

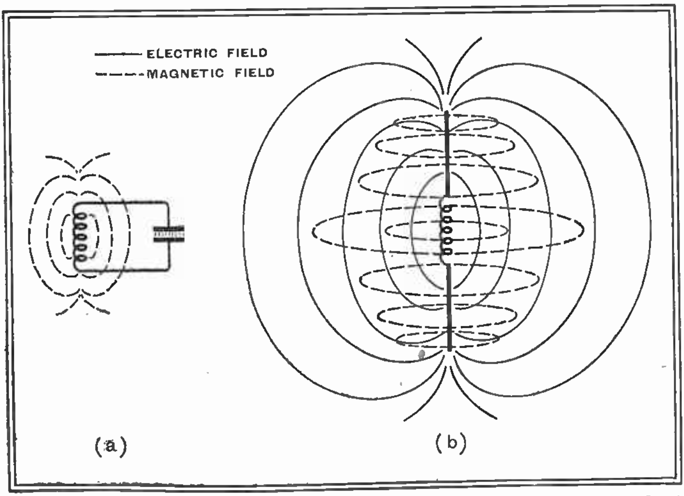


Fig. 162 : In a concentrated tuned circuit (a) the electric and magnetic fields are confined and radiation is slight. Opening out the circuit (b) extends the fields and also the radiation

the condenser? As long as it is concentrated in a small space its radiating powers are almost negligible (Fig. 162*a*). But opening it out so that the lines of force form large loops, comparable with the wavelength, the vertical wires are carrying current in the same direction in all parts of them; and there is no return path to neutralise the resulting radiation. This arrangement (Fig. 162*b*) is a particularly efficient radiator; and it is actually used in large numbers in the form shown, although for convenience the coil is often separated from it as we shall see.

183. The Condenser Radiator

It is not essential for a circuit to be in resonance in order to radiate, but as the radiation is proportional to the current flowing, which is a maximum at resonance, the radiator is practically always tuned. So far we have considered tuned circuits made up of a coil and a condenser as in Fig. 162*a*

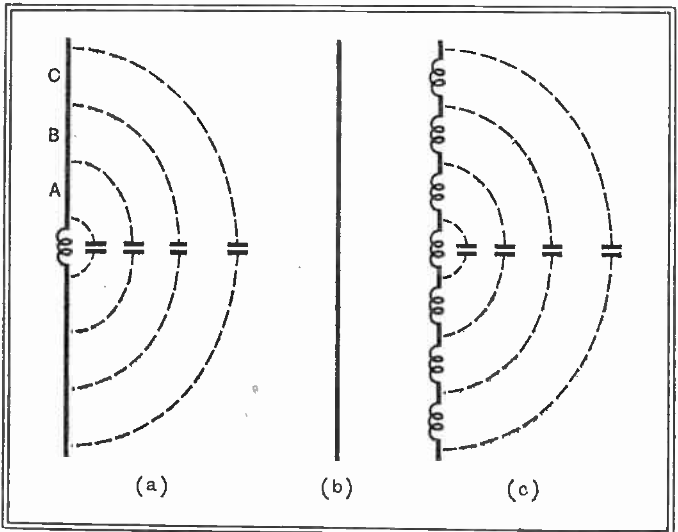


Fig. 163: The distributed capacitance of the radiator in Fig. 162*b* is approximately represented by 163a. A straight wire (*b*) actually has distributed capacitance and inductance, represented approximately by c

These are devices embodying selected quantities of inductance and capacitance in concentrated form. But when the condenser is replaced by long vertical wires as at *b* they possess capacitance which is distributed along their length. It could be approximately represented by a large number of small condensers as in Fig. 163*a*. From this one can see that the current at the point A is greater than that at B, because some of it goes to charge up the section of wire between A and B as well as all that beyond B. Similarly the current at B is greater than that at C, where there is such a small bit of wire to charge up that it requires very little current indeed.

184. The Dipole

As the wires carry this charging current to and fro, pushing an excess of electrons alternately towards the upper and lower halves of the system, they set up a magnetic field ; and the wires, as well as the coil, therefore possess inductance (Sec. 19), which is distributed along them. In fact, a tuned circuit can be constructed entirely of a straight wire (Fig. 163*b*), which can be approximately represented for electrical purposes as a number of tiny inductances and capacitances (*c*). There are, of course, moments in every part of the wire, when the current is zero, when it is just about to turn and come back again ; but, whereas at the extreme ends of the wire it is *always* zero (because there is nothing beyond to charge), at the middle it alternates with the maximum intensity. This condition can be shown by the line marked I in Fig. 164, in which the distance it stands out from the wire at any point represents the R.M.S. or the peak value of the current at that point.

Each half of the wire becomes charged alternately positively and negatively ; but the centre is always midway between these two charges and is therefore at zero potential. The maximum potentials are reached at the ends, for there the charges all along the wire are pushing up behind one another, and the two ends are always at opposite polarity (except instantaneously twice every cycle when they are zero), so the R.M.S. voltage distribution can be indicated by the line V.

As the simple straight wire has both inductance and capacitance, and normally the resistance is comparatively small, it possesses all the qualifications for a resonant circuit, and can oscillate at its own natural frequency. A resonant radiator of this kind is known as a *dipole*. It turns out that the lengths of the resulting waves are almost exactly double the length of wire (actually a few per cent. longer than double), and therefore the arrangement is called a half-wave dipole. For short waves, especially those classified as ultra-short (less than 10 metres), the dimensions of a dipole are so compact that it is a very convenient and popular form of radiator, and can be used for either vertically or horizontally polarised waves, according to the way it is erected.

185. Aerial and Earth

The theory that we have just run through has assumed that the wire is suspended in space, far away from any material substances such as the earth. The conclusions are not appreciably upset by placing it in air, but substances having permeability or permittivity greater than 1 increase the inductance and capacitance respectively and cause an increase in wavelength. The effect due to the necessary supporting insulators is very slight if they are kept small. But the effect of the earth is generally considerable. If the desired wavelength is very short, there is no difficulty in suspending the dipole many wavelengths above the ground; but as soon as radio began to be considered as a means of conveying messages it was found that the very short wavelengths, measured in centimetres rather than metres, have a very restricted range; and a dipole for much longer waves cannot in practice be more than a wavelength or so above ground and the wire itself may be of unwieldly length. It was Marconi who thought of using the ground to form the lower half of a vertical dipole; the remaining half sticking up out of it is therefore often known as the quarter-wave Marconi aerial. The current and voltage distribution are shown in Fig. 165, which may be compared with the top half of Fig. 164.

For the earth to be a perfect substitute for the lower half of the dipole it must be a perfect conductor; which, of course, it never is, though sea-water is a good approximation to it. To overcome the loss due to semi-conducting earth forming a substantial part of the sphere of activity of the oscillating current and its resulting fields, it is a common practice to connect the lower end of the radiator —generally called an aerial at this stage of its development—to a system of radiating copper wires, buried just below the surface to prevent people from tripping over them, shown dotted in Fig. 165. This system is called an earth screen; an alternative consisting of a set of wires stretched just over the ground is known as a counterpoise.

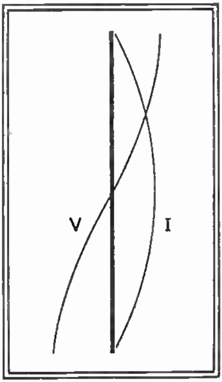


Fig. 164: Distribution of current (I) and voltage (V) in a half-wave resonant radiator

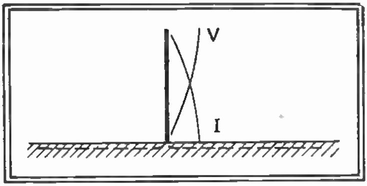


Fig. 165: The earth can be used as one half of a radiator, leaving a quarter-wavelength wire above earth. The dotted line represents buried wires sometimes used to improve the conductivity of the ground

186. Aerial Coupling and Tuning

The next necessity is some means of maintaining oscillations in the aerial. Distributed inductance and capacitance, although satisfactory for composing the aerial itself, are not usually convenient for coupling to a source of oscillations such as a valve oscillator. Oscillatory power can be supplied to the aerial by either capacitive or inductive coupling. The latter is generally the more convenient, and it is therefore necessary to put back some of the concentrated inductance that we removed when arriving at the simple dipole. The result is shown in Fig. 166. The most effective point at which to insert the coil is that at which the current is greatest. The Marconi aerial has a practical advantage here over the complete dipole, for the coil comes near the ground close to the other apparatus instead of awkwardly in mid-air.

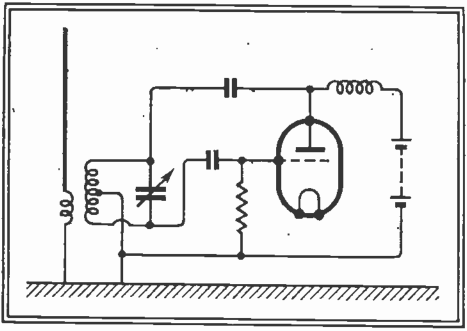


Fig. 166 : Valve oscillator inductively coupled to a vertical aerial

The coil replaces some of the distributed inductance by concentrated inductance, leading to a shorter aerial for a given wavelength and therefore less radiation. To avoid this result, it is possible to neutralise the inductive reactance of the coil by an equal and opposite reactance furnished by a condenser in series with the coil. But in many cases it may actually be desirable to reduce the height of the aerial below its natural quarter of a wavelength : for example it would be quite impracticable to erect a quarter-wave aerial to work on 10,000 metres—it would have to be about 8,000 feet high !—except by means of a barrage balloon, which would be a menace to aircraft and subject to weather conditions. Often, too, it is desired to adjust the tuning by some more handy means than adjusting the length of the aerial wire. It is quite normal practice, then, for a considerable proportion of the tuning reactance of an aerial system, especially for medium and long waves, to be in concentrated form ; but it must be remembered that this is at the expense of radiation efficiency.

Where there are limitations in height, due to restrictions imposed by flying, or by the resources of the owner, it is possible to increase radiation by adding a horizontal portion, giving the familiar T or inverted L aerials (Fig. 167). The effect of the horizontal extension is to localise the bulk of the capacitance in itself so that the current in the vertical part, instead of tailing off to zero, remains at nearly the maximum value and therefore radiates more (Fig. 167*b*). The top portion radiates, too, but with a different polarisation, so the addition due to it may not be very noticeable in a vertical receiving aerial.

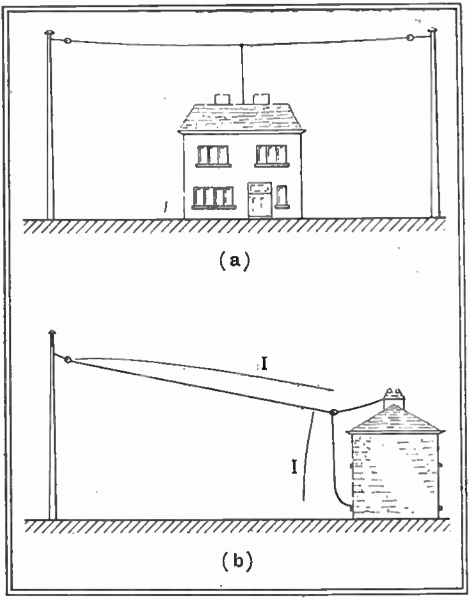


Fig. 167 : To increase the radiation from a vertical aerial without increasing the height, a horizontal " top " is commonly added, forming the " T " (a) or " inverted L " (*b*)

187. Choice of Wavelength

Talking about receiving aerials, the whole of the foregoing principles apply, because the factors that make for efficient radiation are identical with those for the inverse process of deriving the maximum signal power from an electro-magnetic wave.

Before considering receiving aerials in detail, however, it will be as well to know what factors lead to the choice of wavelength. The shorter the wavelength, the smaller and cheaper the aerial and the more efficiently it radiates. For example, a radio link was operated across the English Channel for several years before the War, on a wavelength of 17 centimetres, a half-wave aerial for which is only about

3½ inches long! Why, then, erect at vast expense aerials hundreds of feet high?

One reason is that although great advances are being made in the generation of oscillations at very short waves corresponding to frequencies of 300 Mc/s and more, the

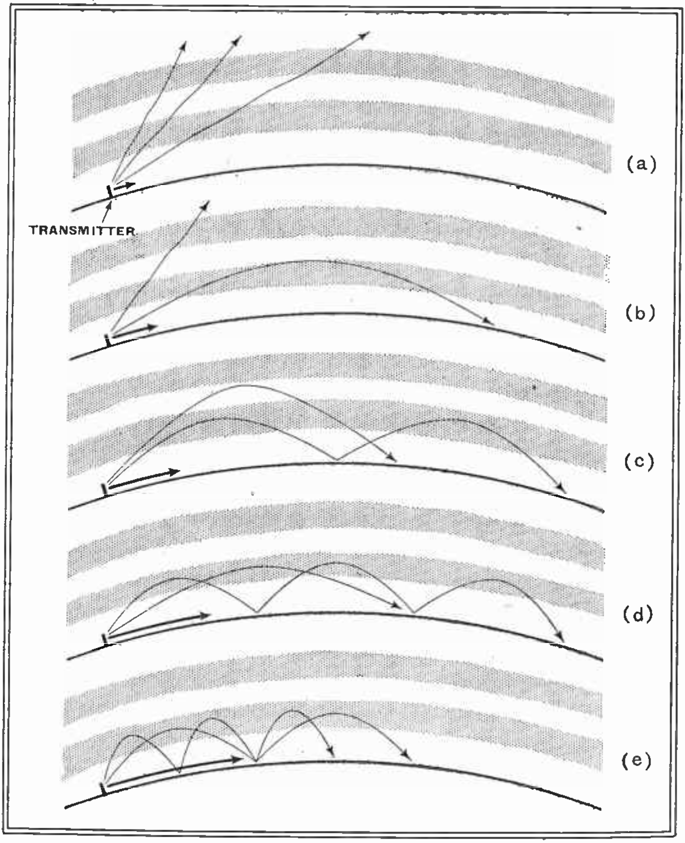


Fig. 168 : Showing (not to scale) the relative ranges of ground wave and reflected wave from very high frequencies (a) to very low (e)

dimensions of the valves and circuits employed are necessarily very small, and their heat dissipation is correspondingly restricted ; there are also other factors that tend to reduce the efficiency ; and so only a comparatively minute amount of power can be handled by them. The other main consideration is that even if plenty of power could be radiated it is fairly quickly absorbed, and not being reflected from the sky it fails to reach far beyond the horizon (Fig. 168*a*). So either the aerials have to be put at the tops of very high towers (in which case their cheapness disappears) or the range is restricted to a few miles. Wavelengths shorter than 10 metres are used for television, not because the shortness of wavelength in itself has any particular merit for that purpose, but because it corresponds to a very high frequency, which is necessary for a carrier wave that has to carry modulation frequencies of up to several Mc/s. They are also used for short-distance communication such as police cars, radiotelephone links between islands and mainland, and other specialised short-range purposes. For a range of even 50 miles a fairly high tower is necessary.

188. Influence of the Atmosphere

As the frequency is reduced below about 30 Mc/s (wavelength greater than 10 metres)—the exact dividing line depends on time of day, year, and solar activity cycle, and other conditions—the range of the wave front travelling along the surface of the earth, and therefore termed the ground wave, increases slowly, while the sky wave is returned to the earth at a very great distance, generally several thousands of miles (Fig. 168*b*). Between the maximum range of the ground wave and the minimum range of the reflected wave there is an extensive gap, called the skip distance, which no appreciable radiation reaches.

As the frequency is further reduced this gap narrows, and earth reflections may cause the journey to be done in several hops (*c*). As the distances at which the sky waves return to earth vary according to time and other conditions as mentioned, it is rather a complicated business deciding which wavelength to adopt at any given moment to reach a certain distance. But a vast amount of data have been accumulated on this and enable fairly reliable communication to be maintained at all times by a judicious choice of wavelength. As waves usually arrive at the receiver by more than one path simultaneously, and tend to interfere with one another, fading and distortion are general unless elaborate methods are adopted for sorting the waves out. At a certain wavelength, of the order of 150 metres, the ground wave and reflected wave begin to overlap *at night*, while during daylight the reflected wave is more or less absent. Over the ranges at which there is overlap the two waves tend to interfere and cause fading and distortion, as they do with more than one reflected wave.

Finally, the range of the ground wave increases and becomes less affected by daylight or darkness, so that waves of 10,000 and more metres have a range of thousands of miles and are not at the mercy of various effects that make long-distance short wave communication unreliable. For this reason they were originally selected as the only feasible wavelengths for long ranges, and are still used for that purpose; but certain serious disadvantages have forced radio engineers to make the best of shorter waves, and now only a small fraction of long distance communication is borne by very long waves. The disadvantages of the latter are (1) the enormous size of aerial needed to radiate them; (2) the low efficiency of radiation even with a large and costly aerial system; (3) the high power needed to cover long ranges, largely due to (4) the great intensity of atmospherics—interference due to thunderstorms and other atmospheric electrical phenomena; and (5) the very limited number of stations that can work without interfering with one another, because the waveband is so narrow in terms of frequency—which is what matters; see Sec. 125.

189. Classification of Frequencies

The following table, which summarises the foregoing, is due to Prof. F. E. Terman:

Class	Frequency range kc/s	Wavelength range : metres	Outstanding characteristics	Principal uses
Low frequency.	Below 100	Over 3,000	Low attenuation at all times of day and year	Long-distance trans-oceanic service requiring continuous operation.
Medium frequency.	100–1,500	3,000–200	Attenuation low at night and high in daytime; greater in summer than winter	Range 100 to 500 kc/s used for marine communication, aeroplane radio, direction finding, etc. Range 150 to 300 and 550 to 1,500 kc/s employed for broadcasting.
Medium high frequency.	1,500–6,000	200–50	Attenuation low at night and moderate in daytime.	Moderate-distance communication of all types.
High frequency.	6,000–30,000	50–10	Transmission depends upon the ionisation in the upper atmosphere, and so varies greatly with the time of day and season. Attenuation extremely small under favourable conditions.	Long-distance communication of all kinds; aeroplane radio.
Very high frequency.	Above 30,000	Below 10	Waves travel in straight lines and are not reflected by ionised layers, so can only travel between points in sight of one another, or nearly so.	Short-distance communication; television; two-way police radio; portable equipment aircraft landing beacons.

190. Beams and Reflectors

Because of the disadvantages, already mentioned, of very long waves, short waves are used wherever possible. A further point in their favour is that it is practicable to concentrate the radiation in any desired direction instead of wasting a large part of the power by indiscriminate distribution. When the wavelength is much greater than the dimensions of the aerial, radiation takes place fairly equally in all directions, and it is difficult to modify this.

But when the aerial is half the wavelength, as in the simple dipole, the radiation varies from nil along its axis, to a maximum all round its " equator " as indicated by the line drawn around the dipole in Fig. 169*a*. The distance of any point on this line from the centre is an indication of the relative strength of radiation in that direction. If a second

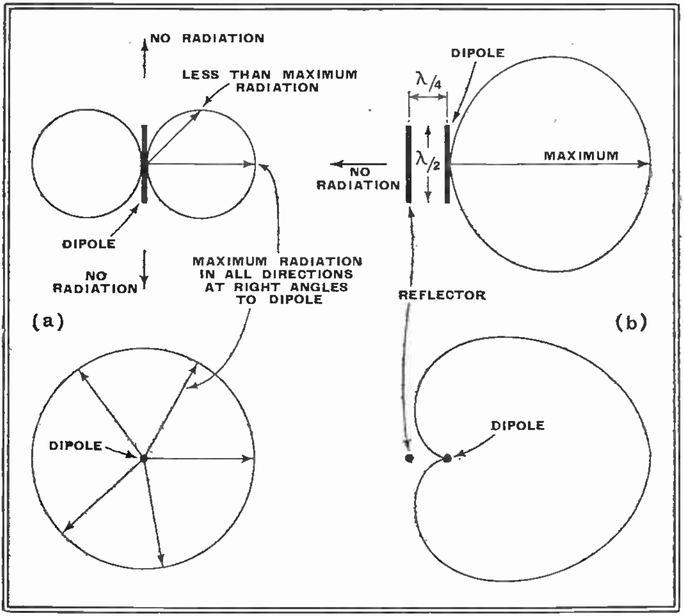


Fig. 169 : Polar diagrams of radiation from a simple dipole aerial (a) ; side view above, end view below, and dipole with reflector (*b*)

half-wave dipole, not fed with power, is placed quarter of a wavelength away, as shown at *b*, it is like a resonant coupled circuit and has oscillatory currents induced in it. These currents re-radiate, and the quarter wave spacing causes the re-radiation to be in such a phase as to cancel out the original radiation on the side where the second dipole is placed, and to reinforce it on the opposite side. The second dipole therefore acts as a *reflector*. The same argument applies to a receiving aerial; and by using reflectors at both transmitting and receiving ends there is a considerable gain in signal strength.

But this is only the beginning of what can be done. By placing a number of dipoles in a row, or end-on, or both, the radiation from them adds up in phase in certain directions, and cancels out in others, and in general the greater the number of dipoles in such an array the narrower and more intense the beam of radiation. Obviously the size of such an array would be impracticable for long waves, whereas for very short waves it is reasonably compact. This compensates to a large extent for the difficulties in generating large amounts of power at very high frequencies; and in fact the short-wave beam system for Empire communications was just in time to prevent a large sum of money being spent on a grandiose scheme for high-power long-wave stations.

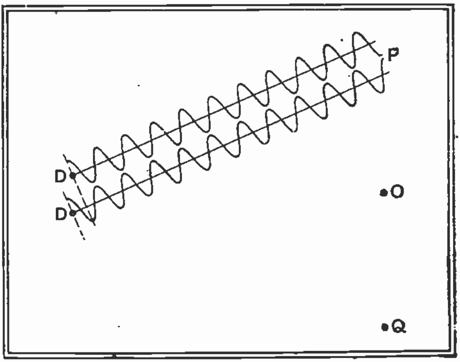


Fig. 170: Showing how the waves from two dipoles radiating in phase arrive at certain angles in completely opposite phase

The principle on which most of the directional aerial arrays, however elaborate, depend can be illustrated by a pair of dipoles shown end-on at DD in Fig. 170. They are fed with R.F. power of the resonant frequency (i.e., wavelength equal to twice the length of the dipoles) *in phase*, so that at any point O equidistant from the two dipoles the radiation from both arrives in phase and gives stronger reception than from one dipole only. But points such as P or Q are at unequal distances from the dipoles, and if the difference in distance happens to equal half a wavelength —or any odd multiple of half a wavelength—the radiation from the two arrives in opposite phase and cancels out. By using a sufficiently large number of dipoles, suitably arranged, the radiation can be concentrated into a narrow angle.

It is necessary to consider, not only the distribution of radiation horizontally around the transmitter, but also the angle of elevation about the horizon. In this the ground has a large influence, and the height of the aerial above ground is an important factor. Obviously if the direct ground wave is being relied upon it is wasteful to direct most of the radiation upwards ; while if reflected waves are necessary to reach the desired destination the upward angle should be adjusted accordingly. The design of an aerial array is therefore far from simple, especially as some of the data—conductivity of earth, and reflecting power of atmosphere, for example—are imperfectly known and subject to change.

191. Direction Finding : Frame Aerials

An important branch of directional wireless is direction finding, commonly abbreviated D.F. It is possible to make use of the fact that dipoles receive nothing when pointing towards the source of radiation to discover the direction in which the source lies. As Fig. 169*a* shows, it may be in either of two opposing directions ; so to distinguish between these a reflector is generally used, as at *b*. If two or more receivers observe the bearing of the transmitter, its position on the map can be determined by drawing lines at the appropriate bearings from the receiving sites and noting where they intersect.

It is more usual, however, to use for directional receiving purposes a coil or frame aerial, which as we saw in Sec. 182 transmits and receives nothing broadside on, and in fact has a characteristic similar to that of Fig. 169*a* but turned through 90°. It has the advantage of being quite easy to tune over a considerable waveband.

The subject of direction finding is a large and complex one, as can be gathered from the fact that Keen's well-known book on it runs to 800 pages, so the details cannot be adequately discussed here ; but it may be noted that directional receiving aerials—notably the frame type—are of value as a means of avoiding interference. No amount of selectivity in the receiver is able to separate two stations working on the *same* frequency ; but if they are situated in different directions from the receiver a frame aerial can be set to cut out the unwanted one. This setting is fairly critical, whereas the strength of reception varies little over a wide range of angle each side of maximum ; so it is possible to discriminate between waves arriving from not very different directions. Another glance at Fig. 169*a* will make this clear. Most portable sets use frame aerials, and as some of this type have very poor selectivity it is just as well !

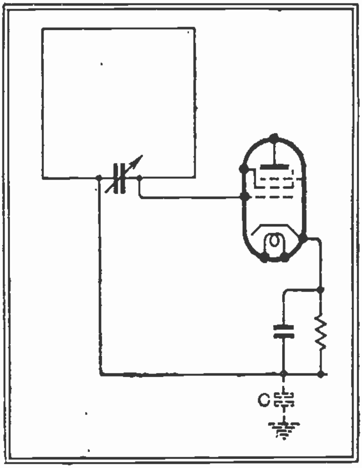


Fig. 171 : Due to its capacitance to earth, a frame aerial is liable to act also as a vertical aerial, receiving a signal even when turned broadside to the source

Unless special precautions are taken, such as electrically balancing the aerial with respect to earth, it will be found that a strong signal cannot be made to vanish at any setting of the frame. The reason for this is that besides acting as a coil it also acts as an ordinary vertical aerial, which has no directional properties. In Fig. 171 a wave striking the frame aerial broadside induces equal voltages in both vertical parts of it, and so far as sending a current *around* the tuned circuit is concerned they cancel out. But besides acting as a coil the frame has a certain amount of capacitance to earth just like any elevated aerial, as indicated by the imaginary condenser C ; and the induced voltages are in parallel for causing currents to flow up and down between aerial and earth. Some of the current sets up a potential difference across the tuning condenser and this is amplified by the receiver. To avoid this the frame must be balanced to earth; and Fig. 172 indicates one of several ways in which this can be done. Potentials *to earth*

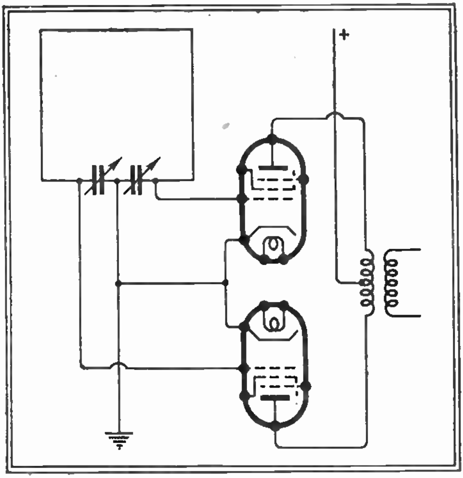


Fig. 172 : One method of avoiding "vertical" effect, by means of a balanced circuit

imposed on the grids of the two valves are equal and of the same polarity, and can be eliminated in the output by a push-pull connection. Potentials resulting from true coil action are in opposite phase and so are amplified.

192. Indoor Aerials and their Disadvantages

The frame aerial is not favoured for transmitting because of its poor radiating properties unless of inconvenient size. Similarly, it is not very efficient as a receiving aerial, and can only be used when either the field strength of the radiation is very great, or a large amount of amplification is employed. This drawback is increased by the fact that it is generally inconvenient to mount a frame aerial out of doors, and indoors it is more or less screened off from signals, while being in an excellent position for picking up undesired noises due to electrical equipment of many kinds—fans, cleaners, razors, etc.—used in houses. The receiver itself also inevitably contributes to noise when amplification is pressed to its limit.

The same applies even more forcibly to what is known as the mains aerial. This consists of a condenser of about 100 $\mu\mu$F connected from the R.F. input of the receiver to one side of the twin mains lead used to supply power to the set. This lead, and the wiring of the house, act as an aerial of a rather inefficient kind. It is most effective, however, in conveying electrical noise to the receiver, originating perhaps from many premises in the district.

A short piece of wire suspended around the room, usually concealed along the picture rail, is little better, for it is closely coupled to the electrical wiring. These indoor aerials, unless used merely as temporary expedients or in situations where an outdoor aerial is impossible, are generally a sign of (*a*) laziness ; (*b*) dislike of the unsightliness of a more conspicuous aerial, or (*c*) a feeling that use of an outdoor aerial indicates possession of a cheap insensitive receiver and hence social inferiority. Many high-class receivers give very poor results because they have such a reserve of amplification that it is considered that "any aerial or no aerial will do ". Actually the important thing is not the strength to which the incoming signal can be amplified—that is generally quite easy—but the ratio between it and "noise." Use of a large efficient outdoor aerial enables the amplification to be cut down and noise reduced.

It is a common fallacy that use of a large aerial reduces selectivity. Actually it enables the coupling between aerial and first tuned circuit to be reduced, which reduces the damping of the circuit and so improves its selectivity, besides making it easier to gang it with other tuned circuits in the receiver (Sec. 97).

193. Anti-Interference Aerials

The increasing use of electrical equipment, much of which causes noise in radio receivers, has led to widespread adoption of anti-interference aerials, the principle of which is to erect the part of the system which is effective for receiving as far away as possible from domestic and industrial sources of electrical noise, and to screen the lead-in and indoors portion, which otherwise is relatively less good for station reception but contributes the greater part of noise pick-up. The higher the aerial proper, the greater the desired signal and the less noise it is likely to pick up. When the level is reached below which any benefit from increased signal reception is likely to be more than outweighed by increase in noise, the whole of the wire from here to the first tuned circuit in the receiver should be completely screened. The receiver circuits themselves are assumed to be screened, and it is necessary to stop the ingress of noise from the mains lead by interposing a R.F. filter—usually a choke coil in series with each wire, and a condenser shunted between them. Whether it is better for the condenser to be on the mains side or receiver side of the chokes depends on local conditions.

Schematically the foregoing arrangement is as shown in Fig. 173. The aerial downlead and lead-in are surrounded by a braided metal sheath, which is earthed. Insulation between wire and sheath is, of course, essential, and consequently the capacitance to earth of the screened portion is very large—far larger than that of the unscreened

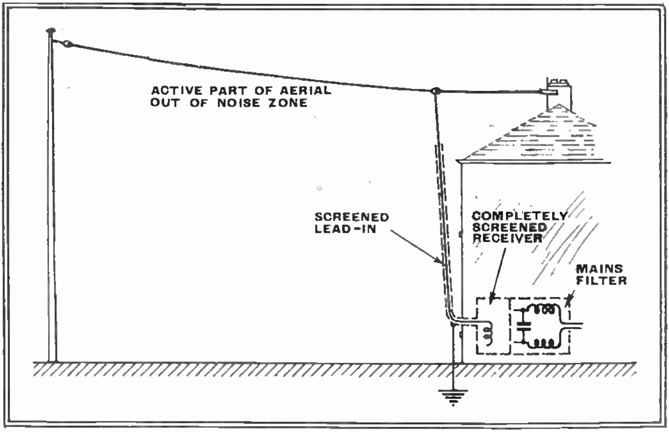


Fig. 173. Diagram of a screened-downlead anti-interference aerial

part. It is like connecting a high-capacitance solid-dielectric condenser across aerial and earth terminals, as in Fig. 174. The dielectric, unless very carefully selected, causes serious R.F. loss (Sec. 61); but, even if it did not, the capacitance shunts a large part of the precious signal current away from the receiver. The higher the frequency, the lower the reactance and the greater the loss. And, of course, the transfer of so much capacitance to the tuned circuit upsets the tuning if, as is usual, it is ganged.

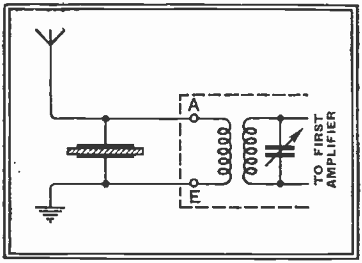


Fig. 174: Approximate electrical equivalent of the screened downlead, causing loss of signal

194. Matching the Impedances

The answer to this problem is in Sec. 48, where the rule was given for calculating the effect on one winding of a transformer caused by connecting a resistance across the other. The same holds good for reactance. Suppose, for example, the capacitance of the screened downlead is 1,600 $\mu\mu$F. At 1 Mc/s that is a reactance of about 100 ohms (Sec. 34). If the transformer between aerial and first tuned circuit were a close-coupled one with a 1 : 1 ratio the effect would be to add 1,600 $\mu\mu$F to the tuning condenser, which would be perhaps 10 times the normal tuning capacitance for that frequency, and the voltage developed across it would be very much reduced. But if it were a 1 : 10 ratio transformer the equivalent reactance across the secondary would be 10^2 or 100 times as much—10,000 ohms, which is the reactance (at 1 Mc/s) of 16 $\mu\mu$F, or only one hundredth part of the capacitance of the aerial downlead. Actually it is quite unnecessary to work out the reactance, for the secondary capacitance is reduced in the same proportion as the reactance is increased—the square of the transformer turns ratio—and, of course, is the same at all frequencies, apart from certain effects neglected by the assumption of an ideal transformer.

There is a limit to the transformer ratio that can be advantageously used in the endeavour to minimise the effect of the large shunt capacitance of the aerial downlead. The primary current increases as the step-up ratio is increased ; and there comes a point where the loss due to this current having to pass through the series impedance of the aerial more than counterbalances any gain due to step-up ratio. It is another case of *matching* the aerial impedance to that of the tuned circuit, as in Sec. 96. The impedance of the type of screened lead generally used is about 80 ohms, regardless of frequency ; whereas that of the tuned circuit varies considerably with frequency, being possibly well over 100,000 ohms at low radio frequencies and 5,000 ohms or less at the very high. So the transformer ratio is a compromise, as we saw in Sec. 97 ; a ratio of from 1 : 10 to 1 : 25 is of the right order except for very high frequencies.

How about the connection between aerial and screened downlead? The impedance at medium and low frequencies of an ordinary non-resonant aerial (as distinct from the dipole or the quarter-wave earthed aerial, which have to be a different length for each wavelength) is also very variable and indeterminate, but it is almost certain to be very much greater than 80 ohms ; and unless there is to be another bad impedance match with consequent loss of signal strength a step-*down* transformer is needed at this point.

The design of such a transformer to cover low, medium, and high radio frequencies is very involved indeed, and compromise reigns almost supreme ; but manufacturers have managed to produce transformers suitable for both top and bottom ends of the screened section (the latter for use where the receiver aerial circuit is not already specially designed for the purpose) in which the whole of the broadcast wavebands are covered ; and, although some loss of signal is inevitable, it is not as great as the reduction in noise is likely to be, and so a net gain in signal-noise ratio is shown.

195. Radiation Resistance

The reason why the impedance of an aerial changes widely as the frequency is varied is that the reactance of its inductance and capacitance both depend on frequency; and it is only when it resonates that the two are equal and opposite, leaving only the resistance to take account of (Sec. 51). In this respect it is like a circuit in which the tuning is not varied to resonate with the applied frequency. If an aerial is required to work on only one frequency, then it is possible to adjust its length to resonate. We have already seen that the dipole type of aerial resonates when it is very nearly half a wavelength long, and the vertical earthed aerial when it is approximately quarter of a wavelength, assuming no "lumped" inductance or capacitance is added. The latter enable aerials of the wrong length to be adjusted to resonance.

When the aerial does resonate, its impedance consists of resistance only, just as in an ordinary tuned circuit. This resistance, when measured at the resonant frequency, is always found to be greater than one would expect from the gauge of wire and known losses. The reason is that, unlike the concentrated form of tuned circuit, it radiates an appreciable part—perhaps a very large part—of the applied radio-frequency power. Although the power is not used up in heating the aerial it is removed from it, and the number of watts radiated as electromagnetic waves is proportional to the square of the current just as is the number of watts lost in heat. The heat loss, W watts, in a resistance R, is I^2R; so by this the resistance is $\frac{W}{I^2}$. If the radiated power is W_r watts, then $\frac{W_r}{I^2}$ can be looked upon as a quantity similar to a resistance, which, when multiplied by the square of the current, gives the number of watts radiated. It is actually called the *radiation resistance*. Obviously an efficient aerial is one in which the radiation resistance is a large proportion of the total resistance, so that most of the applied power is radiated and not much is wasted as heat. Such an aerial is generally good for receiving as well as for transmitting.

A dipole, for example, if made of copper rod or tubing, has an extremely low *loss resistance* ; but owing to its open fields the radiation resistance is quite large, and quickly damps out any oscillation that is started in it without means for maintenance, just like a highly damped oscillatory circuit of the concentrated kind, except that most of the oscillatory power leaves the circuit as radiation and not as heat.

196. Aerials as Resonant Circuits

The resistance of a concentrated tuned circuit is that which would be found by opening it at *any* point and measuring (at the resonant frequency, of course), between the terminals as in Fig. 175*a*. The two reactances—inductive and capacitive—cancel one another out at resonance ; and the resistance normally being small, a

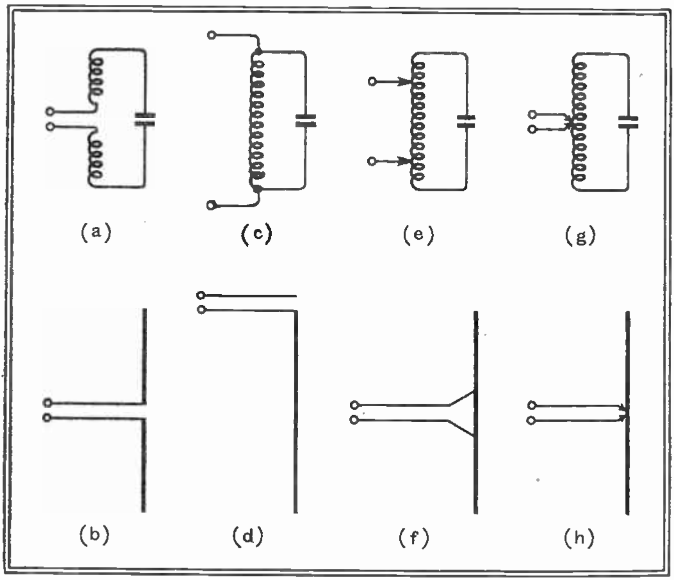


Fig. 175 : The various methods of connecting a dipole, shown in the lower row, are equivalents of the connections, shown in the upper row, to a concentrated tuned circuit

small voltage injected may cause quite a large current to flow. Similarly a dipole can be opened at the centre, where the ratio of current to voltage is large ; and when voltage at the resonant frequency is applied the distributed inductance and capacitance cancel one another out. The loss resistance is relatively small, and the radiation resistance is found to be nearly the same for all resonant dipoles kept far from earth or other bodies and supported by a minimum of insulation. The figure is between 70 and 80 ohms. It may be remembered, incidentally, that this is a normal impedance for a screened aerial downlead, so it is unnecessary to use a transformer for matching one to the other.

When a resonant circuit is connected in parallel (Fig. 175*c*), the currents through the two branches tend to cancel out ; in fact if it were not for resistance it would be unnecessary to feed any current in from outside to keep oscillation going, once started. A normal tuned circuit—resistance small compared with reactance—behaves as a high resistance when connected in parallel (Sec. 56-60), in many practical cases 100,000 ohms or more. The lower the series resistance, the higher the dynamic resistance. Similarly, a dipole, being really a tuned circuit, presents a high resistance across its ends, but, owing to its substantial radiation resistance, its end resistance is not nearly so high as a low-loss circuit ; and, in practice, it is not fed across its two ends, for the leads spaced so far apart would themselves radiate substantially and modify the whole action of the system. A practical alternative is shown at *d*, which looks absurd until it is realised that owing to the distributed nature of the dipole's fields its centre is at zero potential and therefore can be regarded as "earth" even when not so connected. The second wire of the feeder carries a current in opposite phase which partly cancels out the radiation from the first. Used in this way the dipole presents a dynamic resistance of about 3,000 ohms.

Just as the dynamic resistance of a tuned circuit between the connecting leads can be reduced by "tapping down" (*e*), so too can a dipole (*f*). Points can be found presenting a resistance of several hundred ohms, suitable for matching a feeder line of that amount. As we saw in the preceding

Section, and Sec. 48, *e* is equivalent to a step-up transformer. In either case—tuned circuit or dipole—the reactances in both branches between the tapping points cancel out, but they are less in amount and the dynamic resistance is less accordingly. Going to extremes, when the tappings are brought closer together the reactance between them dwindles and finally reaches zero when they coincide (*g* and *h*).

Except when used in push-pull, one end of the coil and of the condenser are at earth potential, and so it is more usual to connect one of the tappings there ; but as the centre of a dipole is the zero potential point it is normally treated as a push-pull circuit, with balanced connecting lines.

When a dipole, or any aerial, is used at other than the resonant frequency, the impedance is not just resistance ; it includes either inductive or capacitive reactance, so in general is larger ; and besides the loss of efficiency due to being out of tune there is a mismatch between it and the connecting feeder.

CHAPTER 20

TRANSMISSION LINES

197. Feeders

ANTI-INTERFERENCE aerials have already given us an example of having to provide some means of connecting an aerial which is up in the air to a receiver indoors and close to the ground. Dipole aerials are another ; and so are the more elaborate systems used for directional purposes. The connecting links are called feeders or transmission lines. It is generally desirable that they shall not themselves radiate or respond to radiation, because that would modify the designed directional effect of the aerial itself, or introduce undesired interference. This object is easily achieved, in principle at least, by placing the go and return wires very close together—the parallel wire feeder—so that the two radiations or receptions nearly cancel out ; or, better still, enclose one lead completely within the other—the coaxial feeder.

In either case, but especially the latter, the closeness of the two leads means a high capacitance between them ; and hence a low impedance. It might appear at first sight that as the impedance of a condenser decreases steadily as the frequency increases, so would the impedance of a feeder. But this line of argument (which we temporarily adopted in Sec. 194) ignores the *inductance* of the feeder, which behaves in the opposite manner ; and it happens that *under certain conditions* the two effects exactly balance one another, and the impedance of a feeder or transmission line is practically the same at all frequencies, and dependent only on the diameters and spacing of the wires or tubes employed. Moreover, under the same conditions this impedance is entirely resistance, and is equal to the resistance of the load connected to the far end.

Beginners usually have difficulty in seeing how the resistance depends on these things and not on the *length*, which one is accustomed to regard as the most important factor in reckoning the resistance of a piece of wire. It seems quite absurd to say now that whether the line is long or short has nothing to do with its resistance. The reason for this paradox is that at very high frequencies the actual resistance of the wire or tube, which is what depends on its length, is almost or entirely negligible in comparison with its inductance and capacitance. The effect of length is generally taken into account in a different way, by saying that it causes a loss of such and such a proportion of the signal voltage per 100 feet (or other unit of length). Obviously a very thin wire, of comparatively high resistance, will cause a greater loss of this kind than low-resistance wire.

To understand exactly what is meant by this new sort of resistance or impedance, which is independent of length, involves mathematics of a higher order than is assumed in this book. But the following explanation, although somewhat crude, may perhaps succeed in conveying some idea of the meaning.

198. Waves along a Line

The inductance and capacitance of a parallel or a coaxial line are distributed uniformly along it, and so cannot be exactly represented by conventional symbols. But Fig. 176, representing a short section of the beginning of a line, is an approximation to it. A high-frequency generator supplies power, and a voltmeter and ammeter measure the R.M.S. voltage and current. Suppose the line is infinitely long, or, if that is too great a stretch of the imagination, so long that we have a little while to examine conditions at the starting end before the wave has had time to travel the whole way to the distant end.

What happens during the first positive half-cycle is that current starts flowing through L_1 and charging up C_1. The inductance prevents the current from rising to its

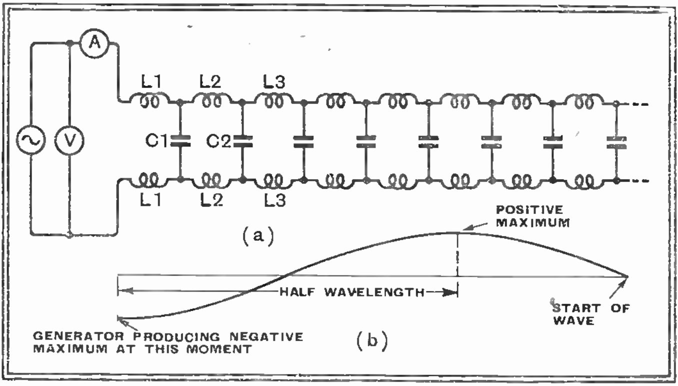


Fig. 176 : A parallel wire line is approximately equivalent to the system shown at a, in which L and C are very small and very numerous. After the generator has turned out three quarters of a cycle the situation is as shown at b

maximum exactly at the same time as the voltage maximum, and it keeps it flowing a little while after the first half cycle is over. In the meantime, the inertia of L_2 to the growth of current through it allows a charge to build up in C_1; but gradually current gets moving in L_2, and so on, passing the positive half-cycle along the line from one section to another. Meanwhile, the generator has gone on to its negative half-cycle, and this follows its predecessor down the line. Assuming the line and its surroundings are of non-magnetic material and are free from solid (or liquid) insulating material (in other words, the permeability and permittivity are both equal to 1) the wave will pass down the line with the speed of light or electromagnetic waves in space. So if, for example, the generator frequency is 50 Mc/s, each cycle lasts for one fifty millionth of a second, and as the wave travels at 300 million metres per second the single wave has gone 300,000,000/50,000,000, or 6 metres down the line. That is, of course, equal to the wavelength. When the current is a maximum at one point along the line it is a maximum in the opposite direction 3 metres either side (Fig. 176*b*).

A mechanical analogy is provided by a long coil of springy wire suspended horizontally. If one end is moved to and fro along the axis, visible waves of compression and expansion of the spring are set in motion along the coil to its far end (Fig. 177). At the moment that one part

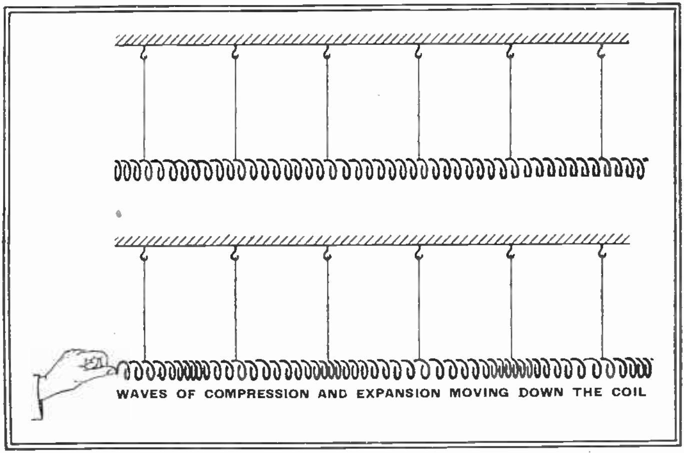


Fig. 177 : Mechanical equivalent of an electrical transmission line. When the "input" is moved rapidly in and out, waves travel along the coiled spring

of the spring is being compressed, parts either side are expanding.

To supply to the starting end of the line the alternating charges that are passed down its length demands from the generator a certain current. Note that the *length* of the line has nothing to do with the strength of current needed, so long at least as the front of the wave has not had time to reach the distant end. Both this current and the voltage that drives it are indicated on the meters, and as the ratio of voltage to current is impedance, we arrive at a figure of impedance which is independent of length and depends only on distributed capacitance and inductance, which in turn depend on spacing and diameter of the wires.

Let us now go, very quickly, to some point along the line and wait for the wave to arrive. If a suitable voltmeter and ammeter are provided at this point we can measure the voltage and current ; and, assuming that the resistance of the wires composing the line, and the leakage between them and the radiation, are all negligible, then the current and voltage are the same as at the start. If there has been no loss on the way, there can have been no reduction in the power sent by the generator.

199. Surge Resistance

Hurrying along to the distant end of the line, we wait for the wave there. Here it comes, the same voltage and current ; until it gets right to the end. Then, if the end is open circuited, there can be no current there. And if it is short circuited there can be no voltage. But suppose we connect a resistance equal to the already measured line impedance, then the ratio of voltage to current exactly satisfies Ohm's Law for that particular resistance, and it absorbs the whole power just as fast as it arrives. As the current going into it and the voltage across it are the same as at the beginning, or indeed any other point on the line, the resistance is equivalent to an infinitely long line of the same impedance. The generator would not know whether the wave started by it was still going on for ever and ever, trying to reach the end, or was being quietly dissipated by a resistance quite a short distance away, *provided that that resistance is equal to the impedance of the line.*

If when the generator supplies 500 volts a current of 1 amp. (alternating, of course), flows into the line, then it is reasonable to say that the impedance of the line is 500 ohms.

In arriving at this result the *length* of the line has not been taken into account at all, for the current starts flowing before the resulting wave has reached the far end. If now a resistance of 500 ohms is connected at that end, when the wave does reach it 1 amp. will flow due to the 500 volts, just as if the resistance were an infinite length of line. This being so, the impedance of the line acts towards the generator as a resistance of 500 ohms, regardless of frequency ; and 500 ohms is called the *characteristic or surge resistance* (or impedance) of the line. The line is said to be properly matched or terminated, and the whole of the power supplied by the generator is delivered to the load resistance at the far end.

In practice, as already stated, a certain percentage of the voltage or power are lost for every 100 feet of line, due to losses by heat or radiation on the way. Apart from this, however, the current and voltage would be exactly the same all along the line, and in fact the line would be a perfect link of any desired length between generator and load resistance, enabling them to be installed in different places without loss on the way. The load being purely resistive, in spite of the reactance of the line, current and voltage are everywhere in phase. There is no tendency for the line to resonate at any frequency, as line inductive and capacitive reactances cancel out.

200. Wave Reflection

If the load is reactive, or is not equal to the characteristic impedance of the line, matters are more complicated. Suppose that in the example considered a load of 2,000 ohms is connected, instead of 500 ohms. According to Ohm's Law it is impossible for 500 volts to be applied across 2,000 ohms and cause a current of 1 amp. to flow. Yet 1 amp. is arriving. What does it do ? Part of it, having nowhere to go, starts back for home. In other words, it is *reflected* by the mismatch or improper termination. The reflected current, travelling in the opposite direction, can be regarded as being opposite in phase to that which is arriving, giving a resultant which is less than 1 amp. In this process a reflected voltage is produced which is *in* phase with that arriving, giving an augmented voltage at the end of the line. If 50 per cent. of the current is reflected, leaving 0·5 amp. to go into the load resistance, then the reflected voltage is also equal to half that arriving, giving in our example a total of 750. A voltage of 750 and current 0·5 amp. would fit a 1,500 ohm load, but not 2,000 ohms ; so the reflected proportions have to be slightly higher—actually 60 per cent., giving 800 volts and 0·4 amp.

We now have 500 volts 1 amp. travelling from generator to load, and 300 volts 0·6 amp. returning, opposite in phase to one another, to the generator. At a distance back along the line equal to quarter of a wavelength the arriving wave and the returning wave are half a wavelength apart (because a return journey has to be made over the quarter wavelength distance). Whatever the voltage or current may be at any point on a wave, half a wavelength farther on it is at all times equal but opposite. So at a point quarter of a wavelength from the load end the reflected current is *in* phase with that arriving, while it is the voltages' turn to oppose one another. The voltage is therefore 200, and the current 1·6 amp. At a point one-eighth of a wavelength from the load and the arriving and reflected currents and voltages are all quarter of a wavelength, or 90°, out of phase, giving 582 volts and 1·16 amps. The same for any odd number of eighth-wavelengths. Half a wavelength along, the arriving and reflected waves are one whole wavelength apart, which brings them into step once more ; so the conditions are the same as at the load.

201. Standing Waves

The currents and voltages can be traced out point by point, and when plotted are as in Fig. 178. It is important to realise that this is not, as it were, a flashlight photograph of the waves travelling along the line ; these are R.M.S. values permanently set up at the points indicated, and would be indicated by meters connected in or across the lines at those points (assuming the meters did not appreciably affect the impedance of the line). Because the arriving and reflected waves, travelling in opposite directions, combine to cause this stationary wavelike distribution of current and voltage along the line, the effect is called *standing waves*. For comparison, the uniform distribution of current and voltage, resulting when the load resistance equals the surge impedance of the line, is shown dotted.

What happens when the reflected wave reaches the generator? It depends on the generator impedance. If it is 500 ohms, and resistive, then the 300 volt 0·6 amp. reflected wave is completely accepted by it, and if the line is an exact whole number of half-wavelengths long the voltage at the terminals of the generator is 500 (outgoing) + 300 (returning) or 800 volts, and the current 1 −0·6, or

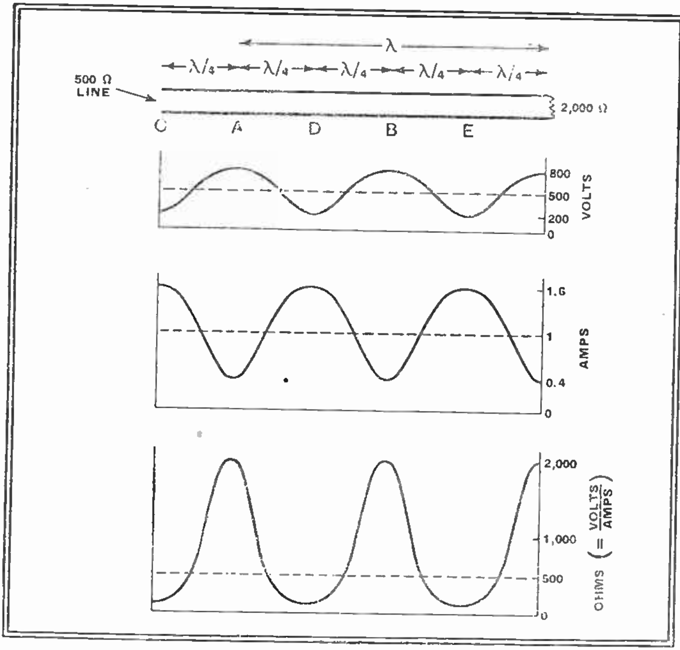


Fig. 178 : The load end of a 500-ohm line, with a 2,000-ohm load, showing the distribution of voltage, current, and (as a result), impedance

0·4 amp. ; in fact, the conditions are the same as at the load end, and the line presents a resistance to the generator given by 800/0·4 = 2,000 ohms, the same as that of the load.

Suppose, however, that the generator resistance is something different, say 2,000 ohms ; then part of the returning wave would be reflected back towards the load, and so on, the reflected quantities being smaller on each successive journey, until finally becoming negligible. The standing waves are the resultant of all these travelling waves, and if a large proportion is reflected at each end—as happens if the impedances there are very much greater or less than the line impedance—it is possible for the voltages and currents at certain intervals along the line to build up to very large amounts. Even though the resistance of the line may be small enough to be negligible when properly matched, the loss due to heating and radiation is proportional to I^2, so may be substantial when standing waves are produced. Quite apart from this, however, the power delivered to a load by a generator—it may be an aerial fed by a valve oscillator—is a maximum when load and generator impedances are matched; and standing waves are generally an indication that they are mismatched, and consequently less than the maximum possible power is being delivered. In any case the line losses are greater than if standing waves are absent.

202. Load-to-Line Mismatch

In the example we have been following through, when the load resistance was equal to the line impedance—500 ohms—the power delivered was $500 \times 1 = 500$ watts. But with a 2,000-ohm load the voltage increased to 800 and the current dropped to 0·4 amp., giving 320 watts. The difference, 180 watts, is represented by the reflected wave, 300 volts 0·6 amp.

It has been stated that the loaded line presents a resistance of 2,000 ohms to the generator *if it is a whole number of half-wavelengths long*. But what if it isn't? To bring it within the scope of Fig. 178, suppose it is $1\frac{1}{4}$ wavelengths long. Then, assuming that the generator resistance is still 500 ohms and therefore the right amount for accepting the whole of the reflected wave on its first return, the voltage at the generator end (as seen from Fig. 178) is 200 and the current 1·6 amp.—again 320 watts. The impedance of the line to the generator is therefore $200/1{\cdot}6 = 125$ ohms.

By selecting this point at which to connect the generator, then, the effect of making the load 4 times the surge impedance of the line is to reduce the load on the generator to one-fourth of the surge impedance. The same applies at any point where the voltage and current have the values stated, that is to say at any odd number of quarter-wavelengths from the load. Although the power delivered by the generator is the same when connected to either a

2,000-ohm point or a 125-ohm point, the power wasted in the generator (I^2R) in the latter event is $1{\cdot}6^2 \times 500 = 1{,}280$ watts, but in the former only $0{\cdot}4^2 \times 500 = 80$ watts. When the load is matched to the line and generator, the power lost in the generator is the same as that delivered to the load—500 watts—so it *may* be considered advantageous to employ a deliberate mismatch in order to obtain an output of 320 watts with a loss of only 80 rather than get 500 and have to lose 500 in the process. In any case, it is obviously better to have too high a load resistance rather than too low (compare Secs. 97 and 139).

From the current and voltage curves it is easy to work out the impedance of a line of any length, and an impedance curve is shown in Fig. 178. It may be noted that at odd numbers of eighth-wavelengths the impedance is equal to the line surge impedance. From a consideration of the travelling currents and voltages in the line it can be shown that only at quarter-wavelength intervals are they exactly in phase, giving a resistive impedance ; at all other points they are reactive as well. To specify the impedance completely, therefore, it would be necessary to analyse it by giving curves of resistance and reactance.

One thing that emerges from all this is that it is possible to do a certain amount of impedance matching by selecting an appropriate point at which to connect. For example, suppose the generator resistance is 2,000 ohms ; then, by connecting it to a point on the mismatched 500-ohm line at which it also presents an impedance of 2,000 ohms (such as A or B), the generator is perfectly matched to the 2,000-ohm load, just as if it were connected straight to it. The only difference resulting from the fact that the surge impedance of the line does not match generator and load is that standing waves are set up on it, which increase the power lost in the line itself. It must be remembered again that although we may be entitled to neglect the loss resistance of the line when calculating its surge impedance, it is usually enough to cause an appreciable loss of power. But if the generator is mismatched to the load it causes the power delivered to the load to be less, even if the loss due to line resistance were absolutely nil.

203. The Quarter-Wave Transformer

A more interesting result is that by shifting the point of connection to one at which the line impedance is 125 ohms (such as C, D, or E), a generator having an internal resistance of 125 ohms can be perfectly matched to a load of 2,000 ohms. The line behaves, in fact, as a 1 : 4 transformer. Similarly, points can be selected giving any ratio between 1 : 4 and 4 : 1, but those marked by letters A to E, are preferable because they avoid complications due to reactance.

It is not necessary to use the whole of a long line as a matching transformer; in fact owing to enhanced losses it is generally undesirable to do so. It can be seen from Fig. 178 that the maximum ratio of transformation, combined with non-reactive impedance at both ends, is given by a section of line only quarter of a wavelength long. It can be seen that the mismatch ratio to the line itself is the same at both ends. In our example the ratio 125 ohms to 500 ohms (generator to line) is equal to 500 ohms to 2,000 ohms (line to load), each being 1 : 4, and, incidentally, equal to the voltage ratio of the whole transformer. If Z_G is the generator impedance, Z_O the surge impedance of the line and Z_L the load impedance, Z_G is to Z_O as Z_O is to Z_L; put otherwise, $\frac{Z_G}{Z_O} = \frac{Z_O}{Z_L}$, so $Z_O^2 = Z_G Z_L$, and $Z_O = \sqrt{Z_G Z_L}$. This formula enables us to work out the surge resistance of the quarter-wave line necessary for matching two unequal impedances. It is often used for linking lines of unequal impedance without causing reflections. In practice it is limited to the fairly narrow ranges of impedance over which it is practicable to construct lines.

That brings us to the calculation of Z_O, the surge impedance. Neglecting the loss resistance of the line, it is equal to $\sqrt{\frac{l}{c}}$, where l and c are the inductance and capacitance per centimetre of the line; and formulæ have been worked out for these in terms of the spacing and diameter of the wires or tubes used. When they are substituted in $Z_O = \sqrt{\frac{l}{c}}$ the result for a parallel wire line (Fig. 179) is $276 \log_{10} \frac{2D}{d}$, and for a coaxial line

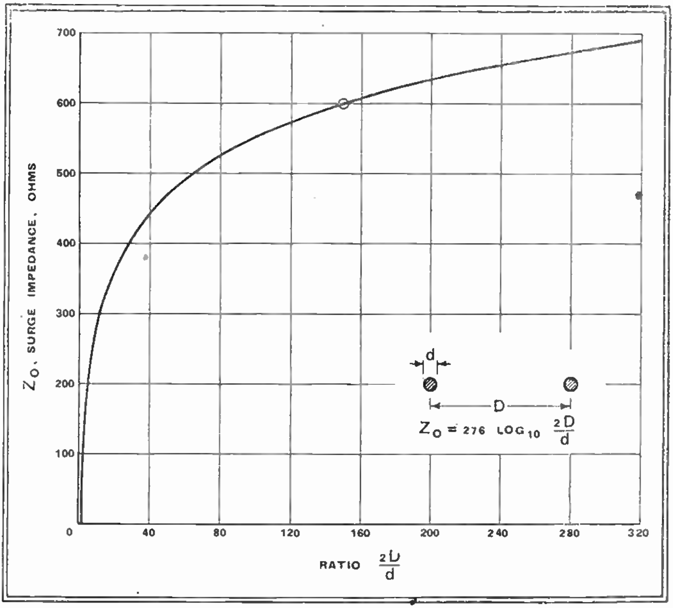


Fig. 179 : Curve giving the surge impedance of a parallel wire line in terms of diameter and spacing of the wires

(Fig. 180) it is $138 \log_{10} \frac{D}{d}$. For reasonable practical values of $\frac{D}{d}$, the surge impedance of a parallel wire line may be 300 to 650 ohms and a coaxial line 60 to 100 ohms. The most efficient proportions give Z_O equal to 600 and 80 respectively. The accompanying curves enable the surge impedance of either type of line to be found from its dimensions, provided that it is air-spaced.

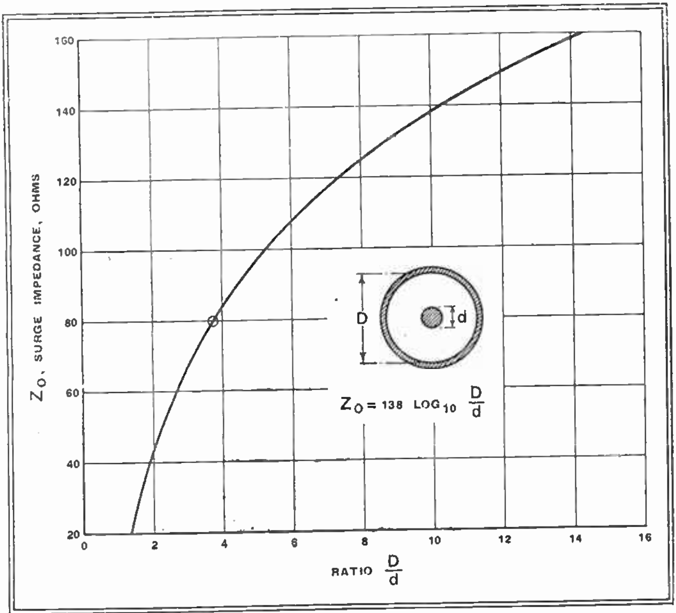


Fig. 180 : Curve similar to Fig. 179, but for coaxial lines

An example of the use of a quarter-wave transformer would occur if it were wished to connect a centre-fed dipole—say 80 ohms—to a 320-ohm parallel-wire feeder. These could be matched by joining them up by a section of line a quarter wavelength long and spaced to give a Z_0 equal to $\sqrt{80 \times 320} = 160$ ohms. A parallel wire line would have to be excess-

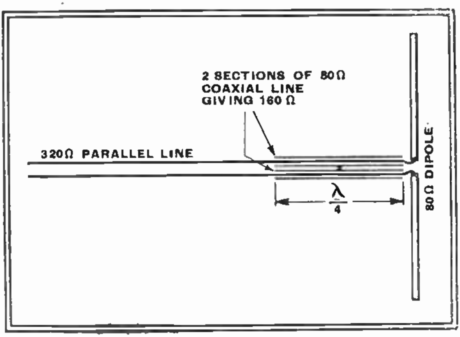


Fig. 181 : One method of matching a dipole to a line of higher impedance

ively close to give this, but the problem could be solved by using a length of 80-ohm coaxial cable for each limb and joining the metal sheaths together as in Fig. 181, putting the impedances in series across the ends of the 320-ohm line.

204. Fully Resonant Lines

Going now to extremes of mismatch, it is of interest to inquire what happens when the "load" resistance is either infinite or zero; in other words, when the end of the line is open-circuited or short-circuited. Take the open circuit first. If this were done to our original 500-ohm example (Fig. 178), the current at the end would obviously be nil, and the voltage would rise to 1,000—double its amount across the matched load. The same would apply at points A and B; while at C, D, and E the voltage would be nil and the current 2 amps. The impedance curve would fluctuate between zero and infinity.

With a short-circuited line, there could be no volts across the end, but the current would be 2 amps.; in fact, exactly as at E with the open line. A shorted line, then, is the same as an open line shifted quarter of a wavelength along. Reflection in both cases is complete, because there is no load resistance to absorb any of the power.

If the generator resistance is very large or very small, *nearly* all the reflected wave will itself be reflected back, and so on, so that if the line is of such a length that the voltage and current maximum points coincide with every reflection, the voltages and currents will build up to high values at those maximum points—dangerously high with a powerful transmitter. These maximum points are called *antinodes*, and the points where current or voltage are zero are *nodes*.

When the length of a short-circuited or open-circuited line is a whole number of quarter-wavelengths, the input impedance is approximately zero or infinity. An *odd* number of quarter wavelengths gives opposites at the ends—infinite resistance if the other end is shorted, and vice versa. An *even* number of quarter wavelengths gives the same at each end.

In between, as there is now no load resistance, the impedance is a pure reactance. At each side of a node or antinode there are opposite reactances—inductive and capacitive. If it is a current node, the reactance at a short distance each side is very large ; if a voltage node, very low. Fig. 182 shows how it varies. It is clear from this

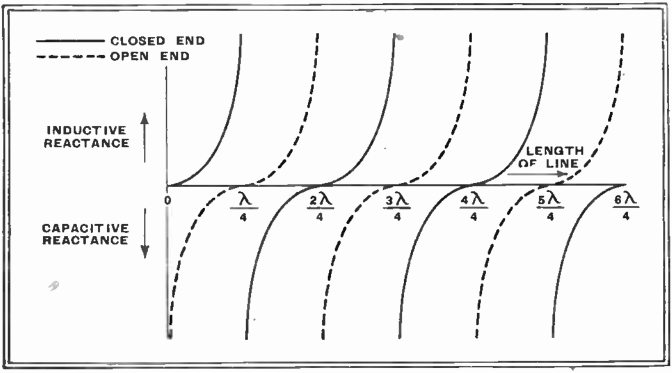


Fig. 182 : Showing how the reactance of a line varies with its length

that a short length of line—less than quarter of a wave-length—can be used to provide any value of inductance or capacitance. For very short wavelengths, this form is generally more convenient than the usual coil or condenser.

205. Lines as Tuned Circuits

It can be used for neutralising reactance of the opposite sort, in matching feeders or bringing them into tune. When the line is quarter of a wavelength long it neutralises its own reactance, and in that respect resembles the ordinary resonant circuit. In fact, it is much to be preferred to the conventional tuned circuit at wavelengths less than about 2 metres. Across its open end it presents a very high resistive impedance—compare the dynamic resistance of a tuned circuit—and to match lower impedances all that is necessary is to tap it down, just as if it were any other sort of tuned circuit.

The parallel wire or bar type lends itself to push-pull

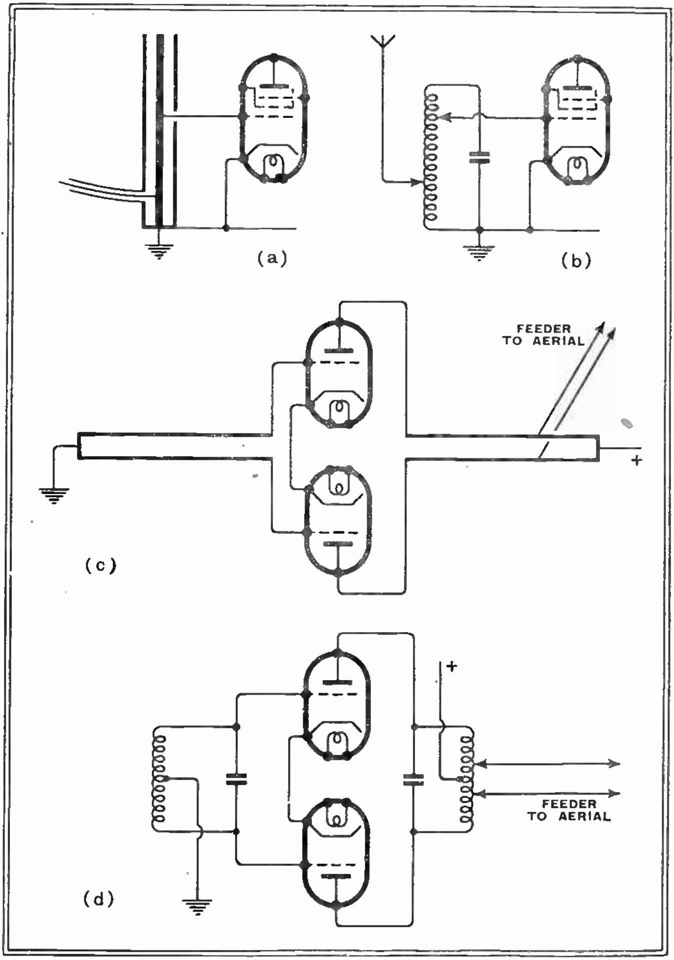


Fig. 183 : Two circuits in which distributed or "line" tuned circuits are substituted for the well-known concentrated type. *a* and *b* are unbalanced, and *c* and *d* balanced or symmetrical

connection, and the coaxial type to single-ended circuits. Owing to their high efficiency, it is easier to obtain oscillation than with coils and condensers. Fig. 183*a* is an example of a coaxial tuning circuit, such as might be used in a receiver, and *b* is the conventional equivalent. A coaxial aerial feeder is used, and being normally about 80 ohms, is tapped low down, near the earthed end. The impedance at the top end is normally many thousands of ohms, and may be too high for the input of a valve, which is quite low at very high frequencies.

Fig. 183*c* is an example of a push-pull oscillator, with its equivalent at *d*.

These are some of the increasingly common applications of transmission lines.

APPENDIX

The reader of various books and articles on radio is very liable to be confused by different terms used to mean the same thing. The following list has therefore been compiled to help clear up the matter. In most cases the first to be mentioned is the one most commonly used in this book. The associated terms are not necessarily *exact* equivalents. Terms distinctively American are printed in italics.

Radio—Wireless

Radio frequency (R.F.)—High frequency (H.F.)

Audio frequency (A.F.)—Low frequency (L.F.)—Speech frequency—Voice frequency

Intermediate frequency (I.F.)—Supersonic frequency

Ultra-high frequency (U.H.F.)—Very high frequency (V.H.F.)

Harmonic—Overtone

Frequency—Periodicity

Root-mean-square (R.M.S.)—Effective—Virtual

Capacitance—Capacity

Permittivity—Dielectric constant—Specific inductive capacity

Q—Magnification

Dynamic resistance—*Antiresonant impedance*

Surge impedance—Characteristic impedance

Feeder—Transmission line

Aerial—*Antenna*

Earth—*Ground*

Frame aerial—*Loop antenna*

Valve—*Vacuum tube*—*Tube*—*Audion*

Anode—*Plate*

Anode A.C. resistance—Anode resistance—Valve impedance—*Plate impedance*

Anode battery—H.T.—" *B* " *battery*

Filament battery—L.T.—" *A* " *battery*

Grid battery—Bias battery—G.B.—" *C* " *battery*

Mutual conductance—*Transconductance*

High-vacuum—Hard

Gas-filled—Soft—Low-vacuum

Tetrode—Screen-grid valve (but not all tetrodes are screen-grid valves)

Heptode—Pentagrid

Auto-bias—Self-bias

Detection—Rectification—*Demodulation* (in U.S.A.: British usage reserves this term for a different phenomenon)

Reaction—Retroaction—Feedback—*Regeneration*

Quality (of reproduced sound)—*Fidelity*

Automatic gain control (A.G.C.)—Automatic volume control (A.V.C.)

Frequency changer—First detector

Reaction coil—*Tickler*

Tuned circuit—*Tank circuit*—LC circuit

Condenser—Capacitor

Coil—Inductor

Screen—*Shield*

Moving coil loud speaker—*Dynamic loud speaker*

Telephones—Phones—Earphones—Headphones—Headset

Radiogramophone—*Combination*

Interference—Jamming

Atmospherics—Strays—X's—*Static*

Noise—Machine interference—*Man-made static*

INDEX

Reference should also be made to the list of Contents on page xi, and alternative terms on pages 322-3

INDEX

INDEX