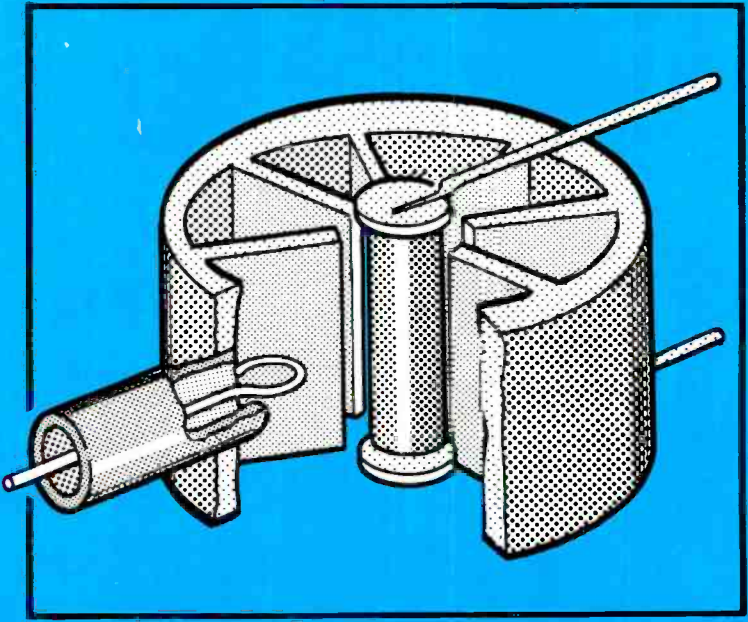
An Introduction to Microwaves

F. A. WILSON



AN INTRODUCTION TO MICROWAVES

Other Books of Interest

BP254	From Atoms to Amperes
BP293	An Introduction to Radio Wave Propagation

AN INTRODUCTION TO MICROWAVES

by

F. A. WILSON

BERNARD BABANI (publishing) LTD
THE GRAMPIANS
SHEPHERDS BUSH ROAD
LONDON W6 7NF
ENGLAND

First Published – January 1992

British Library Cataloguing in Publication Data
Wilson, F. A. (Frederick Arthur)
An introduction to microwaves.
I. Title
621.3813

ISBN 0 85934 257 3

Printed and bound in Great Britain by Cox & Wyman Ltd, Reading

Preface

> *Read not to contradict and confute, nor to believe and take for granted, nor to find talk and discourse, but to weigh and consider.*
>
> Francis Bacon, *Of Studies*

Although microwaves had been around for some time, it was not until 1947 that their use for cooking was introduced. Since then slowly but surely the term has become a household word but how many households have the faintest idea of what a microwave really is? Naturally we of the electronics faith know all about them – or do we? Just in case we are unsure how to answer this question, this book has been written. Its main purpose is to ensure that microwaves are no longer shrouded in mystery for we can no longer afford to be indifferent to their accomplishments since many facets of our society would collapse without them. For example, multitudes of telephone calls travel by microwave, the waves carry our television over land or by satellite, aircraft land safely and even wars are won with their help.

This book is not for the expert but neither is it for the completely uninitiated. It is assumed that the reader has some basic knowledge of electronics, including semiconductors. There is some additional help in the Appendices for readers who have yet to become involved in the field of communications.

Although electronics is often explained with the aid of mathematical equations for they are a concise and accurate way of expressing ideas, they tend to dishearten many would-be enthusiasts. Accordingly the temptation to indulge in detailed discussions has been avoided with long-winded mathematical reasoning giving way to a mention of the final outcome only..

To those of us brought up on wired-in components, the study of microwaves expands our way of thinking into wave theory and the idea of electromagnetic waves propagating through space instead. The use of microwaves grows inexorably, thus creating an ever increasing need for at least some basic understanding of their nature.

Nature may be wonderful, she controls our lives but not man's ingenuity. We read this while surrounded by a profusion of man-made radio waves, unseen and unheard until a device fabricated from the very earth on which we live picks out the one we require. Such is radio, truly a fascinating invention but who can really appreciate, yet alone understand, an electromagnetic wave reversing its direction more than one thousand million times in only one second! Nevertheless as Shakespeare's soothsayer said:

In Nature's infinite book of secrecy
A little I can read.

Read on and for us may the little become a little more.

F. A. Wilson

Contents

Chapter 1

INTRODUCTION

Way back in the late 1800's microwaves made their first appearance when Heinrich Hertz experimented with an oscillating electric spark and was able to demonstrate the existence of radio waves. Unbeknown to him this first transmission happened to contain microwaves. Later at the turn of the century Marconi demonstrated radio communication using "Hertzian waves" by sending the first wireless telegram across the Atlantic. In those early days radio wavelengths were measured in hundreds of metres. Nearing the end of the same century we talk of gamma and X-rays which have wavelengths of less than one hundred thousandth of one millimetre. Somewhere in between are the microwaves to which we are indebted for their part in such impressive developments as television (both terrestrial and satellite), wide band communication systems, radar, heating, and the less agreeable electronic warfare systems. For many the few notes which follow on some of the underlying features of electromagnetic waves may help in clarifying the explanations which follow.

1.1 Charge

Fundamental to everything in this book is the phenomenon of *charge*. This is something Nature has bestowed upon us, never seen, never fully understood but as with gravity we have been able to work out the laws governing it. We can measure it, feel its effects and most certainly use its powers. No material exists without charge which with gravity holds the universe together and is the fundamental driving force of electricity. Charge is in fact defined as a quantity of electricity.

To begin to understand charge we might look at gravity first. This is also an unseen force, surrounding us all the time, waiting to pounce it seems. Hence if we climb a ladder and slip, gravity brings us down to earth with sometimes disastrous results. How difficult it is to appreciate that such a powerful force can exist unseen in the air all around us. We say we are within the *field* of gravity. All gravitational forces are attractive.

Charges on the other hand are of two different kinds. Within an atom the protons which reside in the nucleus are said to have a positive charge, the electrons orbiting around the nucleus carry (or perhaps are) a negative charge. The golden rule is:

"like charges repel, unlike attract"

so the unlike charges of protons and electrons in an atom attract each other and are responsible for keeping the atom together. They balance *exactly*, hence a complete atom exhibits no net charge. That unending power is exerted when many fundamental particles such as electrons get together becomes evident when we pause to consider that it is their charges which provide the force to drive the many tonnes of an electric train.

Of interest is the fact that we measure charge in *coulombs* (after Charles Augustin de Coulomb, a French engineer and physicist). If 6.242×10^{18} elementary charges (e.g. electrons) congregate together, the total charge is equal to one coulomb. The charge of a single electron is therefore $1/(6.242 \times 10^{18}) = 1.602 \times 10^{-19}$ coulombs. If one coulomb of charge passes a given point in one second, one ampere of current is said to flow.

1.2 Waves

Throw a pebble into a pond and the particles of water displaced by the pebble rise in a ring, this has to be for there is nowhere else for them to go. Then gravity pulls them down so displacing some of the water outside of the ring which in its turn is compelled to rise, again to be pulled down by gravity. So the circle of waves goes on widening, driven by the energy given up by the stone. Although the waves travel outwards it is important to note that a cork on the surface of the water bobs up and down but does not travel with the waves, hence showing that there is no advance of the particles of the medium in the wave direction.

A *waveform* is a curve showing the shape of the wave at any given time as plotted on graph paper or by an oscilloscope.

We meet many other types of wave in everyday life, for example, on the sea, in the bath and, of course, sound waves. All of these require a *medium* through which to travel but our main interest lies in the *electromagnetic wave* which is the only one not requiring a medium for its support, in fact it travels best in a vacuum. But more of this later.

In the scientific world waves are generally described in two different ways: (i) by the *frequency*, i.e. the number of complete oscillations or cycles per second [one cycle per second = 1 hertz (Hz)] ; and (ii) by the *wavelength* which is the distance travelled between two successive similar displacements along the direction of propagation. We can convert between (i) and (ii) provided that the velocity of the wave is known. The simple formula linking frequency (f), wavelength (λ) and velocity (v) is:

$$f\lambda = v \qquad \text{i.e. } \lambda = v/f \text{ or } f = v/\lambda\,.$$

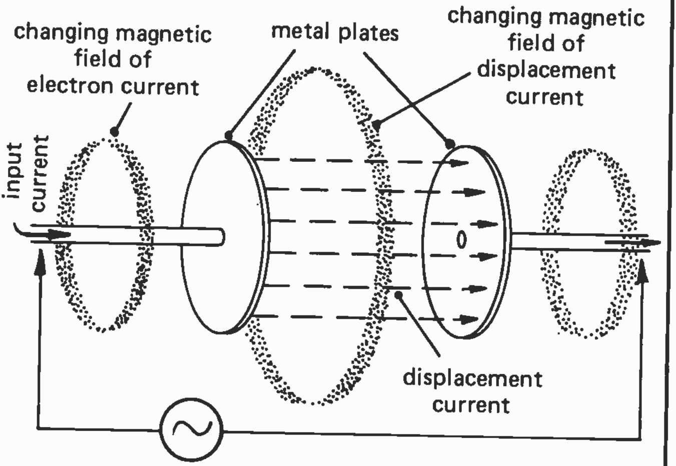
As an example, for a sound wave travelling at 344 metres per second (m/s), a frequency of 1000 Hz has a wavelength of $v/f = 344/1000 = 34.4$ cm.

There seems to be no hard and fast rule as to which description of a wave is used. As frequencies get very high scientists tend to talk in terms of wavelength but at lower frequencies, including microwave, frequency is perhaps more in favour although either frequency or wavelength may be used.

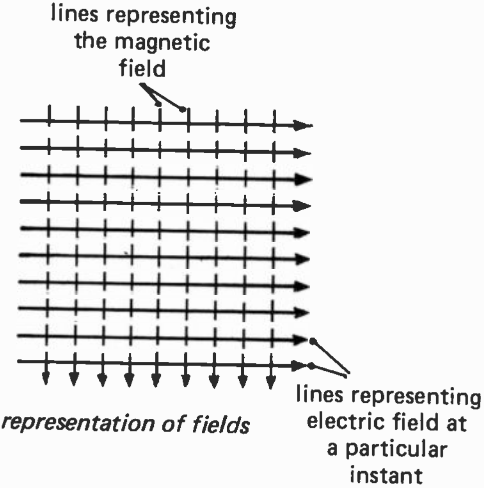
The wave most frequently encountered in electronics is the *sine wave.* It is a pure wave in that it has no components. This is in contrast with all other regularly recurring waves each of which can be shown to be made up of a number of sine waves of differing amplitudes and frequencies.

1.3 The Electromagnetic Wave

Most of us are aware that when a current flows through a conductor a magnetic field is set up around it and also that a changing magnetic field embracing a conductor induces an e.m.f. in the conductor. This establishes a both-way link between electric current and magnetism when a conductor is involved. In fact in the early 1900's Hendrike Anton Lorentz



(i) a displacement current produces its own magnetic field



(ii) simple representation of fields

FIG. 1.1 THE ELECTROMAGNETIC FIELD

was able to show that both current flow and magnetism are simply different aspects of the forces arising from charges in motion. We need to go one stage further however to understand that a changing electric field can give rise to a *displacement current* in an empty space where no conductor is present. In Figure 1.1(i) are two metal plates to which an alternating potential is applied, hence a changing electric field arises between them. James Clerk Maxwell in the 1860's was the first to suggest that a *changing* electric field could produce what he called a displacement current even in empty space and as shown in the figure, that this is just as capable of producing a magnetic field as the current in a conductor. To avoid some heavy-going mathematics, we must accept that displacement current arises from the fact that an applied electric field creates an electric flux through any dielectric, even space and that the current is given by the rate of change of flux with respect to time. It can also be shown to be proportional to the rate of change of the electric field strength so here is a kind of Ohm's Law relationship but referring to changing quantities only.

Using our most convenient way of representing a field, i.e. by arrowed lines, (i) of the figure can be shown as in (ii). The two fields are at right angles.

As the generator in (i) reverses polarity so the arrows on the field lines in (ii) reverse in direction. When the combination of electric and magnetic fields is forced to move outwards we will see from the next section that an electromagnetic wave is produced.

1.4 Electromagnetic Radiation

Electromagnetic radiation occurs when a charge is made to oscillate quickly in anything but we use as an example an antenna wire. A changing electric field ensues which as shown in Section 1.3 sets up a changing magnetic field at right angles to it. In turn the magnetic field creates another changing electric field at right angles but in the opposite direction to the original one. This second electric field produces a further magnetic field and so on. Thus the two fields are linked and they reinforce in such a way as to keep one another going. Perhaps the simplest way of looking at the process is to consider the energy transferred from the oscillatory current to the fields. As the fields collapse this energy would normally be returned to the wire. However, if the charge polarity has reversed before the fields have fully collapsed, all of the field energy cannot be returned because new and opposite fields are in the way. The original fields are therefore forced away from the wire and together they form a train of waves travelling outwards from the antenna at a speed determined by the surrounding medium. For any medium the velocity is given by:

$$\nu = \sqrt{(1/\mu\epsilon)} \text{ metres per second}$$

where μ is the permeability and ϵ the permittivity of the medium.

For free space (i.e. a vacuum containing no nearby objects), the permeability is $4\pi \times 10^{-7}$ and the permittivity 8.854×10^{-12} so using c for the free space velocity, this works out to:

$$c = 2.998 \times 10^8 \text{ or very nearly } 3 \times 10^8 \text{ m/s}$$

which is the figure used generally.

For other media ν is lower because both permeability and permittivity are greater than 1. For air, however, the values are so close to 1 that the velocity is almost identical with c.

We need to be able to quote the directions of the fields of a particular electromagnetic wave, i.e. to state the wave *polarization.* It is only necessary to specify one of the fields since the pair are at right angles. It has generally been agreed to use the electric field for this purpose so a horizontally polarized wave has its electric field lines horizontal (magnetic vertical) and vice versa. Accordingly the wave shown in Figure 1.1(ii) is a horizontal one. The arrows reverse direction at each half-cycle of the wave. A wave may not be truly vertical or horizontal, many things can happen to it in its travels or it may be deliberately radiated at some other angle. Another much used polarization in microwave transmission is *circular.* In this case the electric and magentic fields remain at right angles to each other but now the pair continually rotate. We may need to know which way round they are going so if we look towards an oncoming wave, then a clockwise rotation is described as *positive* or *right-handed* and anticlockwise as *negative* or *left-handed.*

The *plane of polarization* is the plane which contains both the direction of the electric field and the direction of propagation.

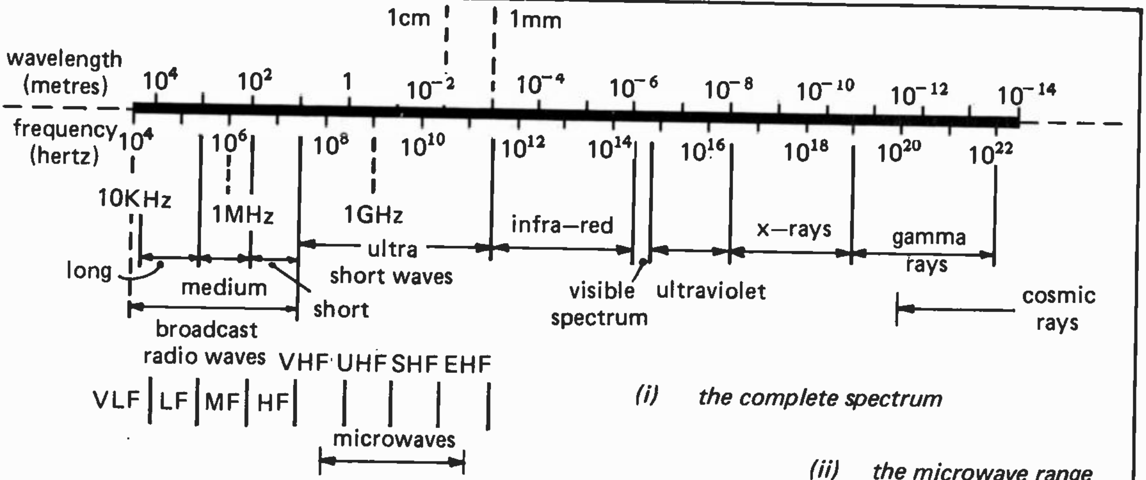
Mode is a term which describes the *form* in which an electromagnetic wave is propagated, more specifically it relates to the electric and magnetic field patterns existing in a waveguide. Many of the patterns are distinguished by code names, these are discussed in greater detail in Section 2.3.

When the fields as shown in Figure 1.1(ii) vibrate at right angles to the direction of propagation, the wave is described as *transverse.* Electromagnetic waves in free space and air are normally transverse and are called TEM waves (transverse electric and magnetic).

1.5 The Electromagnetic Wave Spectrum

In Section 1 is an indication that the range of wavelengths seems to be never-ending, this range is generally called the *electromagnetic spectrum*, running from the longest waves of radio to the tiny wavelengths of gamma and cosmic rays. Within the spectrum are shorter ranges which are classified by their usage. A comparatively short but absolutely essential range is near the centre, these wavelengths give us light, i.e. they are capable of stimulating the light receptors of the eye. Figure 1.2(i) shows at the top the full wave spectrum but it must be appreciated that there are no clear-cut divisions between the various ranges or bands, they merge into one another. Moreover there is little general agreement on the band limits except for the internationally agreed radio-frequency bands shown near the bottom of (i). These are:

Band No.	*Frequency Range*	*Frequency Description*	*Letter Designation*	*Metric Subdivision*
4	3–30kHz	Very Low	VLF	Myriametric Waves
5	30–300kHz	Low	LF	Kilometric Waves
6	300kHz–3MHz	Medium	MF	Hectometric Waves



(i) the complete spectrum

(ii) the microwave range

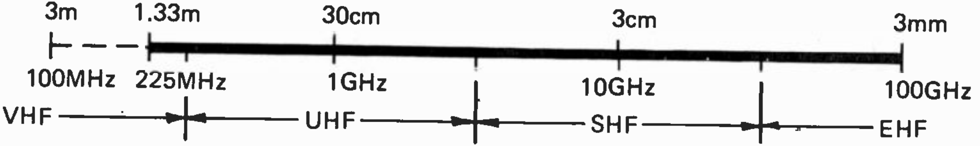


FIG. 1.2 ELECTROMAGNETIC WAVE SPECTRUM

Band No.	*Frequency Range*	*Frequency Description*	*Letter Designation*	*Metric Subdivision*
7	3–30MHz	High	HF	Decametric Waves
8	30–300MHz	Very High	VHF	Metric Waves
9	300MHz–3GHz	Ultra High	UHF	Decimetric Waves
10	3–30GHz	Super High	SHF	Centimetric Waves
11	30–300GHz	Extremely High	EHF	Millimetric Waves

[1 MHz = 10^6 Hz, 1 GHz (gigahertz) = 10^9 Hz]

1.6 Microwaves

The prefix *micro* is not used as elsewhere in electronics to indicate one millionth; coming from the Greek *mikros* it simply means *small*, hence microwaves are those of relatively small (short) wavelength. They are certainly short when compared with long waves of between 1000 and 2000 metres. It is obvious from Figure 1.2(i) that waves considerably smaller exist but microwaves were so named when the potential utilization of wavelengths less than about one metre was only just beginning. At the very bottom of (i) of Figure 1.2 is shown the position of the microwave band and this is amplified in (ii). This is a commonly used one but other ranges than this are to be found and there is appreciable variation. The one shown extends from 1.33 metres down to 3 mm (225 MHz → 100 GHz). As an indication of the degree of variation, here are a few examples taken from modern publications:

50 cm → 1 mm, 100 cm → 1 mm, 30 cm → 0.03 mm,
less than 3 m, less than 20 cm, less than 1 cm.

The band we have chosen is subdivided with letter designations as follows:

P band	1.33 m	→	76.9 cm
L band	76.9 m	→	19.3 cm
S band	19.3 m	→	5.77 cm
X band	5.77 m	→	2.75 cm
K band	2.75 cm	→	8.34 mm
Q band	8.34 cm	→	6.52 mm
V band	6.52 cm	→	5.36 mm
W band	5.36 cm	→	3.00 mm

There is also a C band running from 7.69 cm to 4.84 cm.

These bands are again subdivided and are frequently quoted in radar and satellite technology although again, there is considerable variation, especially with the latter.

Some major developments in microwave technology have been:

(i) the discovery that above about 1.5 GHz microwaves behave in many respects as do light waves. Hence these microwaves can be reflected as can those of light;

(ii) by such reflections, when the wavelength is sufficiently short the wave can be transmitted in a hollow metal tube or *waveguide*. The wave can be pictured as bouncing from wall to wall along the guide;

(iii) the development in the late 1930's of the *klystron*, a device capable of generating a continuous supply of microwaves. This development advanced microwave technology considerably;

(iv) at very short wavelengths microwaves exhibit properties similar to those of infra-red radiation which as Figure 1.2 shows is next door in the spectrum. Infra-red rays are those which produce heat so it was found that microwaves could also be used for heating.

Also it is clear that microwaves have the great advantage over waves of lower frequency in the width of frequency spectrum they offer for communication purposes. From the range shown in Figure 1.2, if we take, say 30 cm to 3mm as an example, some 99 GHz bandwidth is there, i.e. 99 times that already provided by all of the lower broadcast frequencies together which because of the many requirements thrust upon them are very crowded indeed (bandwidth is considered in more detail in Appendix 3). Another, perhaps naive way of looking at this is to consider that if the modulation frequencies occupy the same *percentage* of the carrier frequency in all cases (hardly true but useful for demonstrating the point), then the bandwidth which can be carried at 10 GHz is 100 times that at 100 MHz. The question immediately arises as to why not go still higher in frequency, i.e. into the infra-red region, but then losses when propagating through the atmosphere become excessive (we know only too well what happens to light in a mist or fog). So for broadcast and for point-to-point communication, microwaves have much to offer.

It would have been gratifying to have included in microwave achievements their use in medical electronics. Unfortunately a cursory examination of present techniques shows that at present ultrasonics (sound frequencies above about 20 kHz), X-rays and nuclear rays are in the greatest demand, uses for microwaves are few, one of interest however is mentioned in Section 7.5.

1.7 Managing Microwaves

Circuit techniques used for equipment working at frequencies below the microwave range are unlikely to be successful with microwaves for several reasons, one obvious one being that of regeneration causing instability. Adjacent wires of conventional wiring having only a single picofarad of capacitance between them are in fact coupled by a capacitive reactance as low as 16 ohms at 10 GHz, unwanted energy is therefore easily transferred. Also wire and component inductances develop large unwanted voltages which similarly create havoc. In addition all electric circuits carrying an alternating current radiate energy with a radiated field strength which varies directly as the frequency. At normal broadcast wavelengths this is low enough to be of little consequence but at microwave frequencies radiation from metal parts such as wires and components is appreciable. Such losses cannot normally be tolerated within equipments hence for microwaves different techniques have been, and still are being developed.

Much use is made in microwave technology of transmission lines, coaxial cables and waveguides. Short-circuited lengths of transmission line are also used to provide reactances. With the drive for miniaturization microstrip transmission lines of width 4 mm or less have come into use. For such a small component a relationship with the open-wire transmission line or coaxial cable may seem tenuous but at such short wavelengths this is the order of dimension required. We might sum this up by saying that for microwaves the lumped constant approach generally effective for circuit analysis at the lower frequencies is replaced by electromagnetic wave theory in which transmission lines play a major part.

Advanced semiconductor technology now forms the basis of many microwave devices although magnetrons, klystrons and travelling-wave tubes still hold considerable sway; these are considered in detail in Chapter 3.

An interesting microwave device is the *ferrite* which, to take a single example, when biased by a suitable magnetic field is capable of phase-shifting a microwave and hence can be used as an isolator in that the wave travels in one direction through a device but not in the reverse direction.

These few examples of devices and techniques show that not only are we capable of handling the very high frequencies of microwaves but that much of the experience gained is leading to the development of devices for special applications hitherto unheard of.

Chapter 2

MOVING MICROWAVES AROUND

Any attempt to get to grips with microwave generation and transmission must come to naught unless we have some appreciation of transmission lines and waveguides. Both of these are subjects on their own and as might be expected, full analysis becomes highly mathematical and is therefore not for us here. All we need is to become acquainted with just a few of the elementary principles so that we can form pictures in the mind of how energy travels along lines as waves and then to realize that at microwave frequencies just any transmission line will not do, it has to be related to the length of the wave. We will also find that for waveguides the normal losses due to conductor resistance, insulation losses and radiation are insignificant.

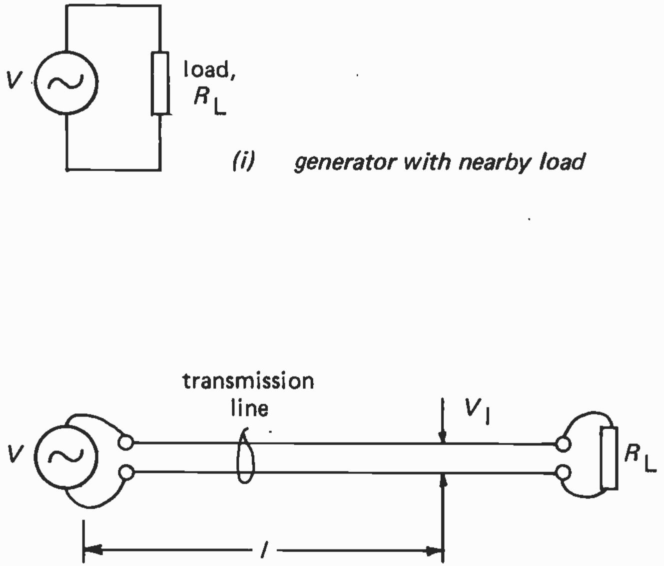
2.1 Transmission Lines

Firstly when does a circuit become a transmission line? Figure 2.1 shows at (i) a generator with its nearby load – no problems here because Ohm's Law is effective with no strings attached. At (ii) the generator and load are separated, it may be over a distance of many kilometres or at very high frequencies a few millimetres. Electrical energy moves between generator and load in a form of wave motion and it would seem that both the time it takes to reach the load and the distance the load is from the generator are involved. We can get some proof of this by accepting the mathematical formula for the voltage V_1 developed across the line at a distance l from the generator. Suppose that the generator applies a voltage V to the line, then:

$$V_1 = V \cos \omega(t + \beta l/\omega)$$

where t is the time of travel, $\omega = 2\pi f$ and $\beta = 2\pi/\lambda$.

Looking more closely at the term $\beta l/\omega$, this can be reduced to l/v where v is the velocity of signal propagation, hence it is clear that length besides time and velocity control the value of



(i) *generator with nearby load*

(ii) *generator with a remote load*

FIG. 2.1 CONNECTING A GENERATOR TO A LOAD

V_l. It is perhaps possible now to appreciate the main differences between "normal" circuit design and that for microwaves. With the normal circuit it does not matter electrically how large or small a resistor is provided that its characteristics are those required, similarly with capacitors and inductors. On the other hand, because as we have seen in Section 1.7 that conventional techniques cannot be used with microwaves, it is clear that at these frequencies dimensions become important and must often be related to the wavelength. This frequently happens to be a useful feature for at such short wavelengths even transmission lines may be small enough to be treated as discrete components.

For those whose transmission line theory has seen better days, the next two sections contain some useful reminders.

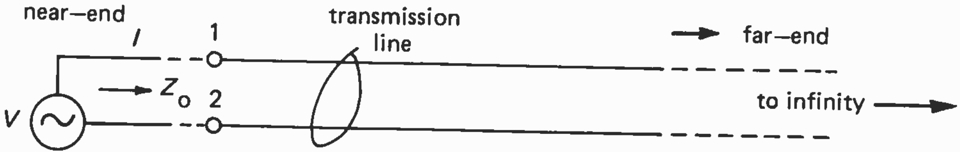
2.1.1 Characteristic Impedance

First let us imagine a line consisting of two conductors disappearing off to infinity. Application of a signal to the conductors (or transmission line) would result in a wave travelling along the line and going on for ever, the line simply acting as a resistance because it can continuously absorb power. This in fact defines the characteristic impedance of a transmission line, i.e. the impedance looking into an infinite length of a line. It is normally designated by Z_0. Figure 2.2 shows at (i) that Z_0 is determined straightforwardly from the values of V and I.

Next suppose that a short length is cut off at the near-end of the line as in (ii) so creating terminals 3 – 6. Looking down the line from terminals 5 and 6 the impedance is still Z_0 because the infinite nature of the line is unaffected, therefore by connecting an impedance Z_0 to terminals 3 and 4 as in (iii) the impedance looking into terminals 1 and 2 is identical with that in (i), that is, the short length of transmission line in (iii) behaves as though it were infinitely long. This makes some sense except that we could not have had an infinite line on which to measure Z_0 in the first place, it can however be shown that certain measurements can be made on a short length of line by which Z_0 can be determined.

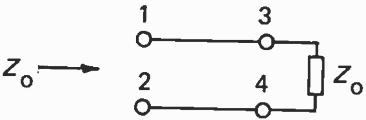
Looking more closely at a short length of the line, it can be seen to have four *primary coefficients* (i) the resistance of the conductors (R), (ii) inductance (L – even a straight wire has inductance), (iii) capacitance between the two conductors (C) and (iv) there is a loss due to the conductance of the insulating material between the conductors, the coefficient is known as the shunt conductance (G). Usually these four values are quoted per metre of line length. Standard transmission line theory arrives at the conclusion that:

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$



(i) measurement

(ii) short length cut off at near-end



(iii) short length terminated in Z_o

FIG. 2.2 CHARACTERISTIC IMPEDANCE

where $\omega = 2\pi f$ and f is the frequency under consideration. j is the standard "operator" used in circuit analysis to indicate a phase difference of 90°.

In practical microwave circuits it is frequently possible to assume that a line is lossless, i.e. that R and G are both zero whereupon Z_0 is more simply expressed as $\sqrt{(L/C)}$.

2.1.2 Reflection and Standing Waves

Reflection arises on a transmission line whenever it is terminated in an impedance other than its characteristic impedance. The Maximum Power Transfer Theorem tells us that maximum power is obtained from a generator only into a matched load. Hence if a transmission line is terminated in its characteristic impedance (Z_0) all the power flowing down the line will be absorbed by the termination. If the termination has a value other than Z_0, say Z_R, then the power cannot be completely absorbed. That which is not absorbed has to go somewhere and the only course open to it is to flow back along the line. This is *reflection* and it is illustrated by the three examples of Figure 2.3 in which in (i) and (ii) at the termination I_i is the incident current and I_r the reflected current. In (iii) for clarity only the line voltage is shown.

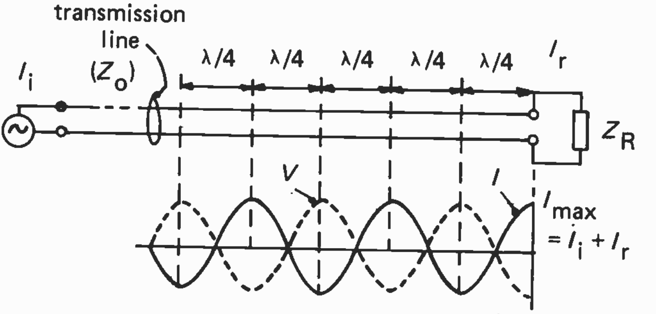
By Ohm's Law and other network theorems it can be shown that:

$$\frac{I_r}{I_i} = \frac{Z_R - Z_0}{Z_R + Z_0}$$

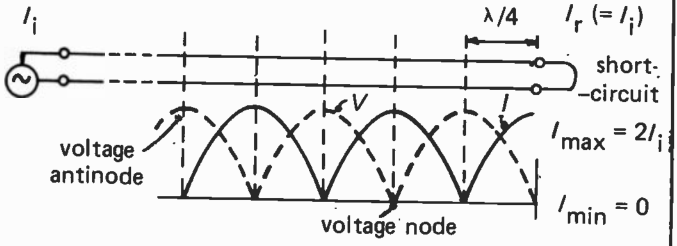
where Z_R represents the impedance at the termination and here we ignore minus signs which arise when Z_R is less than Z_0.

This gives the ratio between the reflected and incident currents in terms of the characteristic and terminating impedances. The ratio also applies to voltages.

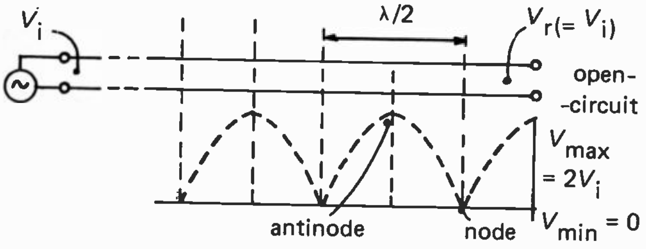
$$\frac{(Z_R - Z_0)}{(Z_R + Z_0)}$$



(i) mismatched line ($Z_R \neq Z_O$)



(ii) short-circuited line ($Z_R = O$)



(iii) open-circuited line ($Z_R = \infty$)

FIG. 2.3
STANDING WAVES ON A TRANSMISSION LINE

is called the *reflection coefficient* (symbol r, but ρ may also be used).

The graphs in the figure show line voltages and currents and as in all cases of mismatch, the current and voltage maxima and minima occur at the same points along the line. The waves shown are known as *standing waves*, having points of minimum voltage with maximum current and vice versa. The points of zero voltage are known as *nodes*, of maximum voltage they are called *antinodes.* The nodes are exactly half a wavelength apart as are the antinodes. To us this is a most interesting feature of a mismatched transmission line for irrespective of the power sent down the line, when the distant end is short-circuited or open-circuited, there are points half a wavelength apart where no voltage (or current) can be measured even though both incident and reflected waves are travelling past. We can see also from the figure that the impedance of the line varies from zero when the voltage is zero but the current is maximum to infinite when the opposite applies with appropriate values in between. This is extremely useful for it allows us to cut short lengths of transmission line to load or transform impedances, with due regard to frequency of course (see Sect.2.1.3).

The ratio of maximum to minimum voltage on a transmission line carrying a standing wave is known as the *standing wave ratio.* We usually refer to the voltage and call it the *voltage standing wave ratio* (vswr). The maximum value (V_{max}) of a standing wave in terms of the reflection coefficient (r) follows from:

$$V_{\mathrm{max}} = V_{\mathrm{i}} + V_{\mathrm{r}} = V_{\mathrm{i}}(1+r) \quad \text{since } r = V_{\mathrm{r}}/V_{\mathrm{i}}$$

also the minimum value:

$$V_{\mathrm{min}} = V_{\mathrm{i}} - V_{\mathrm{r}} = V_{\mathrm{i}}(1-r)$$

hence:

$$\mathrm{vswr} = \frac{V_{\mathrm{max}}}{V_{\mathrm{min}}} = \frac{1+r}{1-r}$$

from which:

$$\text{reflection coefficient}, r = \frac{\text{vswr} - 1}{\text{vswr} + 1} .$$

The vswr is an extremely useful parameter for use in the analysis of microwave circuits, especially in that once its value is known many other transmission line relationships can be calculated from it.

2.1.3 Matching

Frequently it is found that a transmission line load does not have an impedance equal to the characteristic impedance of the line itself, a typical example of this is an antenna which has an appreciable reactance and is connected to a main feeder of resistive characteristic impedance. It is important to reduce or eliminate resulting reflections and therefore high vswr's on the feeder. For these and many other mismatched conditions it is possible to add or interpose a transmission line section to match the two impedances or perhaps merely to neutralize or "tune out" antenna reactance. This is the process of matching to appreciate which we must first look more closely at how the input impedance of a mismatched transmission line varies with line length.

Figure 2.3 shows how the voltage and current along a line are dependent solely on the *load* impedance and it is this which determines their values at the point of connection. At any point on the line therefore the impedance is simply V/I but note that there is a phase difference between them. This highlights one of the important features of mismatched transmission lines which is that the impedance conditions for the whole line are determined at the far-end of the line and that the impedance across the line is repeated along the line at half-wavelength ($\lambda/2$) intervals. The input impedance of a line an exact number of half-wavelengths long is therefore always equal to the load impedance irrespective of the line characteristic impedance. However, this repetitive condition is clearly frequency dependent for it can only occur at those frequencies at which a particular length of line is an exact number of half-wavelengths long.

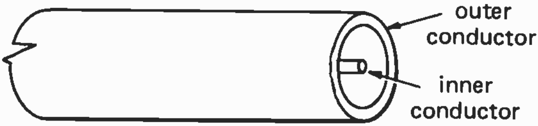
Moreover, it can be shown that because of the phase differences between V and I, quarter-wavelength sections of line have capacitive or inductive reactances depending on whether they are open or short-circuited at the far-end.

All the features of the mismatched transmission line above have considerable usage in microwave technology for at such frequencies, wavelengths are short enough for transmission lines with mismatched terminations to be practicable. As an example, at a medium wave broadcast wavelength of 500 m a quarter-wavelength line is 125 metres long, whereas at, say, 1 cm it is a mere 2.5 mm. Both have the same electrical characteristics at their chosen frequencies but there is little doubt as to which is the more functional. Again we see how special techniques are available with microwaves.

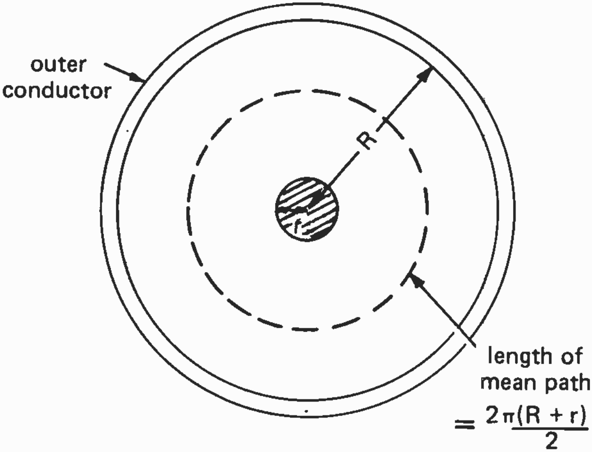
2.1.4 Resonant Cavities

Standing waves are also the basis of *cavity resonators.* If a waveguide which is one half of a wavelength long is closed at both ends it then becomes a closed structure bounded by conducting walls. If a signal at the appropriate frequency is injected into it, a standing wave pattern is generated and in fact the cavity can then be considered as the microwave equivalent of a high-Q parallel resonant circuit. For lumped constant resonant circuits (i.e. inductor and capacitor) reactance is required but resistance is not because it creates power losses. The quality factor or Q is the ratio between them. At microwave frequencies, however, it is better to consider Q as the ratio of the energy stored in a device to the energy dissipated over a specified time interval (e.g. one cycle of oscillation). In practical resonators the energy dissipated per cycle is small compared with the energy stored and typically Q may be many thousands when the cavity is unloaded but when couplings are added (see Sect.2.3.2), the Q falls to a few hundreds. It is lower if a dielectric other than air is used but increases as the conductivity of the metal walls increases. The simplest of cavities is a cylinder with its diameter equal to the length so that a single half-wave pattern is set up in each of the three directions.

If the length of transmission line (coaxial or waveguide) is l, then resonance of the cavity occurs when:



(i) type in common use



(ii) path of non-TEM modes

FIG. 2.4 COAXIAL CABLE

$$l = n \times \lambda_g/2$$

where n is an integer. λ_g is the wavelength in the resonator, not the wavelength in free space. It can be calculated from the free space wavelength, λ, the cut-off wavelength, λ_{co} and the relative permittivity of the dielectric (if not air).

Many other resonator shapes and modes are employed as is seen later.

2.2 Coaxial Cables

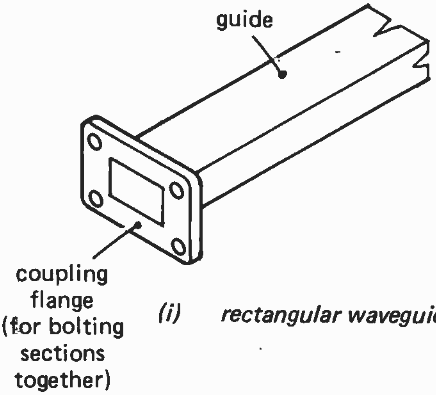
An extremely useful type of transmission line for use with microwaves is the *coaxial cable.* The type in general use has one conductor shielding the other as sketched in Figure 2.4(i), the outer conductor being woven copper braid with the centre conductor kept in place, for example by polyethylene discs. Cable size naturally depends on its power handling capacity.

The characteristic impedance can be calculated by substitution for L and C in the approximate formula $Z_0 = \sqrt{(L/C)}$ and this results in:

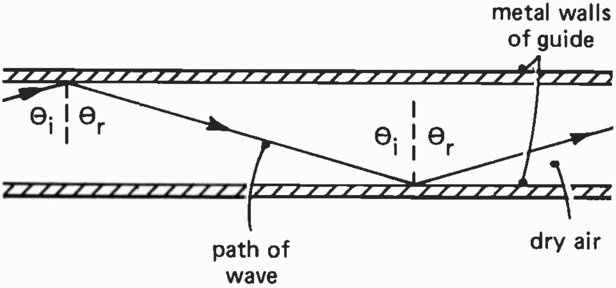
$$Z_0 \approx 138 \log_{10}(R/r)$$

in which R is the inside radius of the outer conductor, r is the radius of the inner conductor and it is assumed that the permittivity of the dielectric (mostly dry air) = 1. See (ii) of Figure 2.4. Generally values for Z_0 range between 50 and 100 ohms non-reactive.

It is important to avoid transmitting at a wavelength which allows additional unwanted modes to propagate. We can imagine from Figure 2.3 that a length of transmission line one wavelength long turned around on itself so that the end is connected to the beginning (i.e. in a circle) is in fact an open-circuited line and therefore is capable of sustaining standing waves, if provided with a small amount of energy from the main wave. This is a condition which can arise within the dielectric of a coaxial tube, the length of the unwanted transmission line being equal to the mean of the circumferences of the inner and outer conductors as shown in (ii) of the figure, i.e. $2\pi \times (R + r)/2$. Not only can this condition arise for a path length of λ, but also for multiples of it.



(i) rectangular waveguide



θ_i = angle of incidence

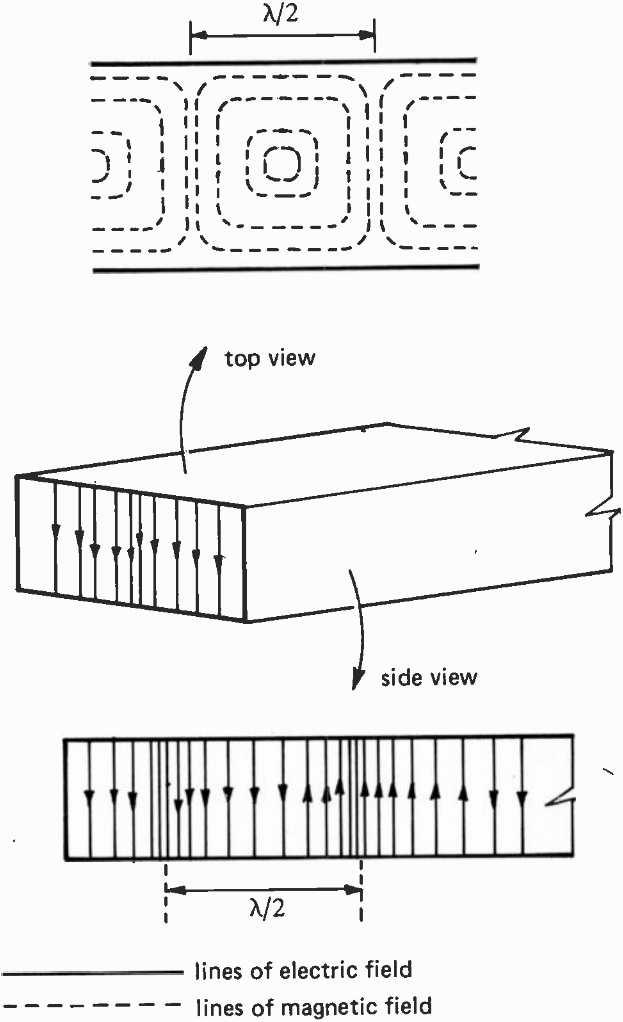
θ_r = angle of reflection

$\theta_i = \theta_r$

(ii) zig-zag path of wave in a guide

FIG. 2.5 WAVEGUIDES

continued



(iii) fields in a TE_{10} mode

continued

Accordingly a TEM wave (Sect.1.4) being transmitted along a coaxial cable will encounter interference and suffer loss if such parasitic non-TEM modes are set up in the dielectric, generally therefore cables are operated below the frequency at which:

$$\lambda_c = \frac{\pi(R + r)}{n}$$

where λ_c is the *cut-off wavelength* and $n = 1, 2, 3, \ldots$ This limits the use of coaxial cable to around 3 GHz (10 cm) above which waveguides take over.

2.3 Waveguides

Omitting the centre conductor of a coaxial cable gives the circular *waveguide.* This is a tubular metal guide in which the electromagnetic wave is confined rather than propagating freely in space. Usually the dielectric is dry air. However, waveguides are more frequently found with rectangular cross-section, an example is shown in Figure 2.5(i).

2.3.1 Guiding the Wave

James Clerk Maxwell first developed the electromagnetic theory of light which then led directly to the discovery of radio waves. Some of his work relates to the *boundary conditions* which are present when electric and magnetic fields travel through one material bounded by another. This may be better appreciated if we recall the principles of optical reflection from our schooldays, remembering that light and microwaves are both electromagnetic waves differing only in wavelength, hence they observe the same laws of reflection at suitable conducting surfaces. Briefly, when light enters a more dense medium its velocity falls resulting in a change of angle of the wavefront at the surface (except at 90°). This is *refraction* and as the angle of incidence (with the normal) of the wave increases, there is a point at which the wave cannot be refracted but undergoes *total internal reflection*, so for example, with glass the surface behaves as a perfect mirror.

The same principles apply to microwaves and under total internal reflection the angle of reflection of the wave is equal to the angle of incidence as shown in Figure 2.5(ii). This is how an electromagnetic wave is propagated through a waveguide, by continual reflections from the guide walls as shown in the figure (which makes it very complicated when we try to picture in our minds the progression of a wave). There is a phase reversal at each reflection. For the electric field in a waveguide there is a change in the permittivity (ϵ) at the boundary or wall. Similarly for the magnetic field there is a change in the permeability (μ). Now an electric field running parallel to a conductor will in fact be short-circuited, hence electric lines of force must always be at an angle to the guide walls. On the other hand the magnetic field will create surface (skin effect) currents in the guide walls, these currents determine the waveguide loss.

2.3.2 Propagation Modes

Things are less complicated if here we ignore the circular waveguide and concentrate on the rectangular, the basic principles are the same for both. We now realize that because it is an electromagnetic wave which is being propagated it is essential to ensure that transmission is entirely through the dielectric and not along the walls of the guide. To accomplish this various *modes* of propagation are used, all ensuring that the wave progresses in a zig-zag fashion from wall to opposite wall as indicated in the foregoing section. The mode is identified in the specification of the type of wave. As an example, for a rectangular waveguide the mode is described by TE_{mn} which indicates that the electric field is transverse to the direction of propagation (transverse electric) with m maxima (or half-wavelengths) occurring across the larger dimension and n maxima along the smaller dimension. Of most importance are the *TE* and *TM* (transverse magnetic) modes, these are always at right angles to the direction of propagation. The simplest mode is the TE_{10} and pictorially this can be shown for a particular instant as in Figure 2.5(iii).

For an appreciation of practical waveguide sizes here are some approximate dimensions for the TE_{10} mode with a dry air dielectric:

to carry 3 GHz ($\lambda = 10$ cm)	7.5 cm × 4 cm
to carry 12 GHz ($\lambda = 2.5$ cm)	2.2 cm × 1.2 cm
to carry 100 GHz ($\lambda = 3$ mm)	4.0 mm × 2.5 mm.

Each transmission mode has a lower limit to the frequency which can be propagated. For any transmission line there is a *propagation coefficient* which expresses the rate at which the current falls as the wave progresses along the line or in this case, along the waveguide. Clearly this is a complex quantity and for a lossless transmission line the propagation coefficient must be purely imaginary for there to be no attenuation, only changes in phase. This is not practicable for as we see in Section 2.3.1 there must be some loss. Nevertheless for the TE_{10} mode, working from the formula for the propagation coefficient, it can be shown that the frequency of cut-off for any particular rectangular waveguide of internal dimensions x (larger) and y (smaller):

$$f_{co} = \frac{c}{2} \sqrt{(m/x)^2 + (n/y)^2}$$

where c is the velocity of electromagnetic waves in space.

This is for an air dielectric, for any other the formula becomes more complicated because the permittivity and permeability must also be taken into account. For the TE_{10} mode $m = 1$, $n = 0$, therefore with air dielectric, theoretically:

$$f_{co} = c/2x \qquad \therefore \lambda_{co} = 2x$$

and in practical rectangular guides this is found to be approximately so but remember this is for the TE_{10} mode only.

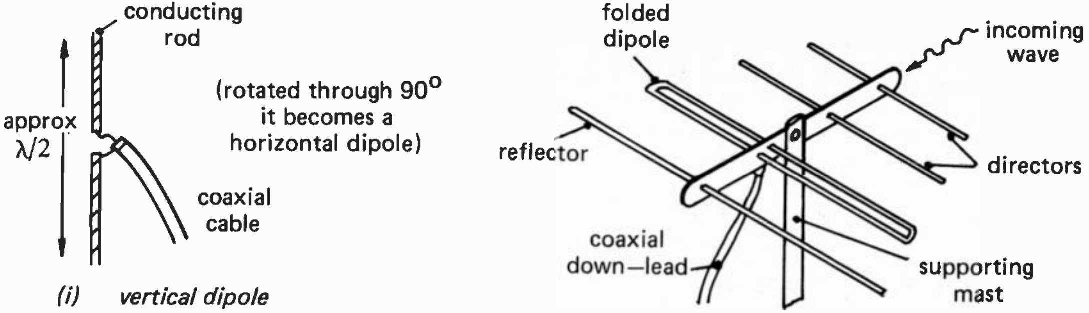
An electromagnetic wave can be inserted into a waveguide using either a voltage driven probe (e.g. by extending the centre conductor of a coaxial cable into the end of the guide) or by a current carrying coil. The desired mode of transmission depends on the alignment of the probe or coil within the guide. A wave is extracted from a waveguide by a similar arrangement. Certain microwave oscillators may be coupled to a waveguide by a special design of mount so that energy is transferred and in the correct mode.

2.4 Antennas

When transmitted by coaxial cable or waveguide the microwave signal may be regarded as captive and it is destined to emerge at the end of the line only. More generally transmission is via a radio path in the atmosphere or space in which the wave is less constrained or may even be broadcast. For radio transmission there are the inevitable antennas (aerials), one for transmitting into the radio path and one or more for receiving from it. There is a multitude of different antenna systems each depending on frequency, requirement and the inventiveness of the designer. Even for microwaves the subject is wide-ranging and includes such types as *pyramidal horn, spiral* and *slotted.* However, we can usefully reduce the range to two types only, both of which are more easily understood and in fact are those most commonly used. These are the *dipole*, confined mainly to the UHF band (Fig.1.2) and the *parabolic* which takes over at frequencies above this. Generally these antennas are reciprocal in that their basic properties are similar on both transmit and receive, also they can be designed to be highly directional.

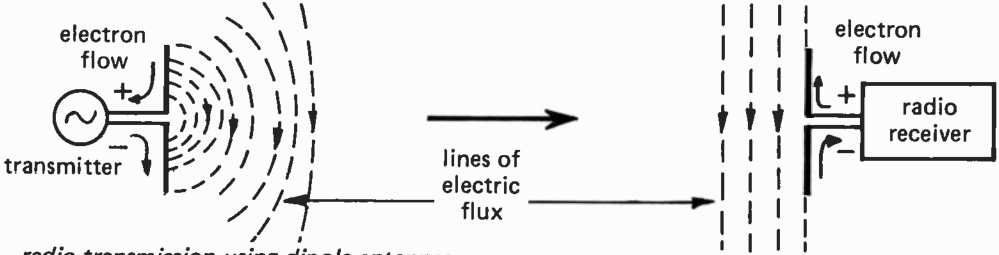
2.4.1 Dipoles

Any piece of wire carrying a radio frequency current radiates energy into the surrounding space but rather inefficiently, antenna design aims simply to increase the efficiency of radiation with added directivity if required. Fundamentally there is a constant relationship between the physical dimensions of the antenna and the wavelength used. The simplest design is the half-wave ($\lambda/2$) dipole, an arrangement consisting merely of a wire or rod split at the centre so that it has two poles into which radiation currents flow as sketched in Figure 2.6(i). Actually the length of the dipole is slightly less than the physical half-wavelength in free space (of the order of 5%), this is because the velocity of the electromagnetic wave in the rods is slower than it is in space. Typcially a half-wave dipole will have a length of approximately $142.5/f$ metres where f is the frequency in MHz. This indicates the suitability of this type of antenna at the lower microwave frequencies for at say, $f = 500$ MHz the dipole length is only 28.5 cm.



(i) vertical dipole

(iii) a 4—element Yagi receiving antenna



(ii) radio transmission using dipole antennas

FIG. 2.6 DIPOLE ANTENNAS

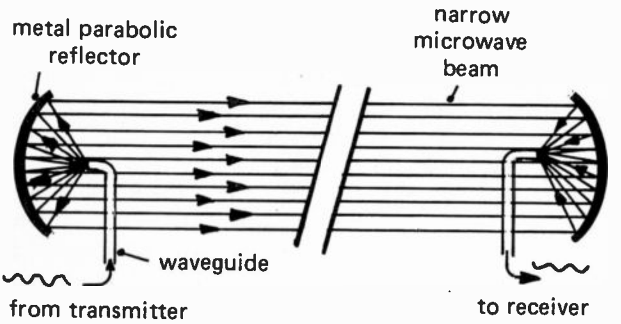
Recalling that the lines of an electric field are arrowed according to the direction in which a free positive charge would move, (ii) of the figure shows pictorially the sequence of events between a transmitter and receiver at a particular instant. Only the electric field is shown, for clarity the magnetic part of the electromagnetic wave is omitted.

Much can be done to enhance the performance of the elementary receiving dipole especially by the *Yagi* system (after two Japanese engineers, H. Yagi and S. Uda) which increases the gain and improves directivity. Yagi antennas have proliferated on our roofs since UHF television transmissions began and one has only to look at them to see whether the electromagnetic wave for the particular area is horizontally or vertically polarized. Figure 2.6(iii) shows a 4-element array with one reflector and two directors (many more are frequently used). These extra rods are excited by the oncoming wave and their spacings from the dipole are such that they re-radiate energy to arrive at the dipole in the desired phase. A *folded dipole* is shown, this has the effect of increasing the bandwidth and raising the antenna impedance.

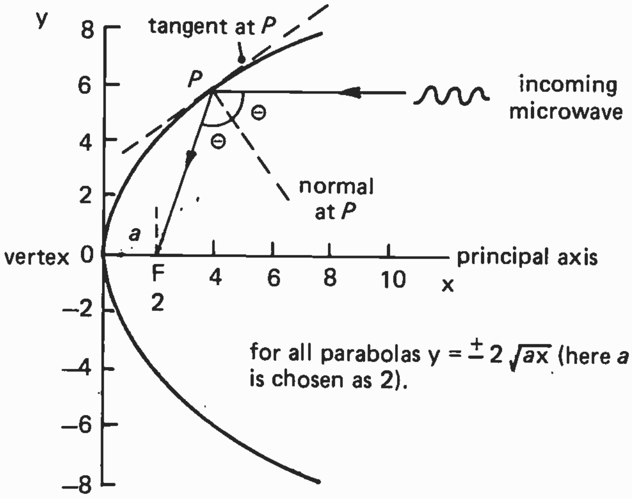
For transmitters, dipole systems may also be used, usually several dipoles are connected together in parallel in a *stacked array*, spaced and fed so that their outputs are all in phase.

2.4.2 The Focused Wave

We are reminded in Section 1.6 that frequencies in the SHF range have wavelengths so short that in many respects they can be handled as for light waves. Just as a searchlight can focus a powerful light at its centre into a narrow parallel beam so equally a *parabolic* antenna focuses a microwave signal into a narrow beam. Practically all of the power is concentrated in the beam and none is lost elsewhere. At the receiving end a similar parabolic antenna receives the beam and focuses it onto a central waveguide. This is illustrated in Figure 2.7(i) with (ii) reminding us of the basic features of the parabola. The shape is that seen when a cone is sliced parallel to its side and we ought to be conversant with the generation of a parabolic curve because it features so prominently in microwave radio systems.



(i) unidirectional point-to-point microwave radio transmission



for all parabolas $y = \pm 2\sqrt{ax}$ (here *a* is chosen as 2).

(ii) essential features of the parabola

FIG. 2.7 THE PARABOLIC ANTENNA

Referring to (ii) of the figure, it can be shown that for any electromagnetic wave arriving parallel to the principal axis, as with light reflected by a mirror, the angle of incidence with the normal is equal to the angle of reflection (θ in the drawing). By repeating the construction shown as P moves around the curve it can be demonstrated that all parts of the wave are reflected towards the *focus F*. F is chosen in the design process and is situated on the principal axis at a distance a from the vertex. Equally a wave generated at F will be reflected along a path parallel to the axis. However, not all microwave dish-type antennas are of the true parabolic shape. A practical difficulty with the parabolic is that the horn-like device which must be located at the focus for delivering the electromagnetic wave in the case of a transmitting antenna or for collecting the wave on a receiving antenna, is itself within the path of the free wave and therefore to some extent reduces the effective area of the dish. Accordingly designs are available in which either the wave is focused to a lower point so that the device is no longer in the way or a second reflector is added to guide the wave to the centre of the main dish. In fact now that satellite television is well established, designs of all shapes and sizes can be seen.

2.5 Ferrites

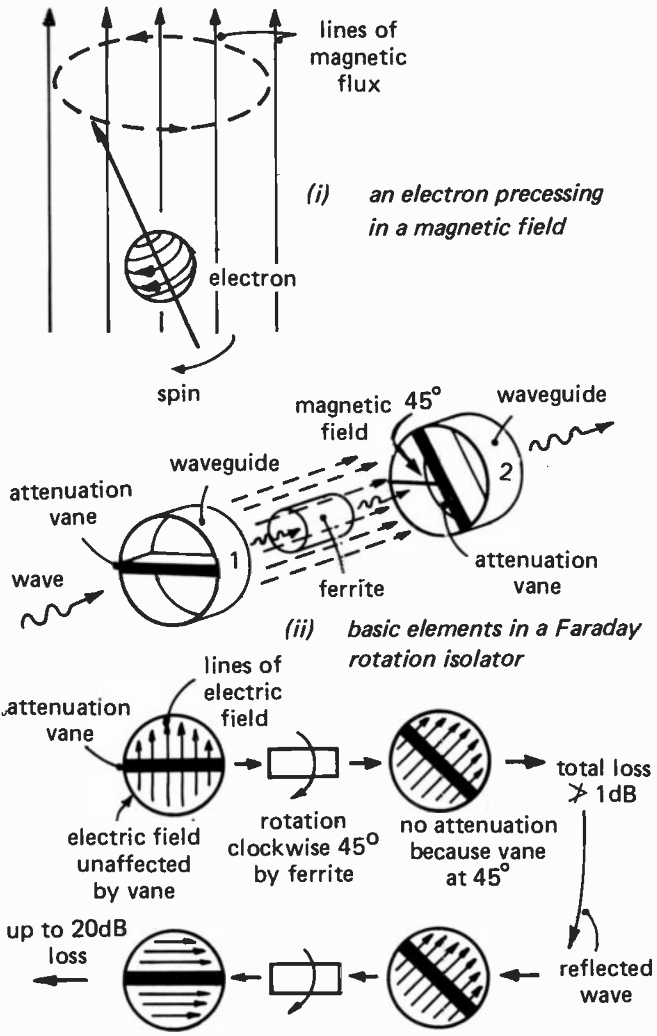
Here our interest lies in the fact that ferrites are devices through which an electromagnetic wave can propagate but in doing so the wave polarization is rotated. There are many instances in microwave transmission where such a device is called for. Ferrites are based on ceramic (pottery-like) materials.

An electron spins rather like a top and behaves as a tiny magnet. In atoms containing paired electrons with the spin of one balanced by the spin of its partner there is no net spin hence the atom exhibits no overall magnetization. On the other hand, in metals such as iron, nickel and chromium each atom has a single unpaired electron so the atom itself is equivalent to a small magnetic dipole (i.e. has N and S poles). If in a material all the magnetic dipoles are aligned into the same direction, that material exhibits strong magnetism. Conversely the dipoles may be aligned in opposite directions but because they may not all be of the same magnetic strength, there is overall a weak magnetism. The material is said to be *ferrimagnetic* and ferrites are of this type.

For microwave applications certain oxides are preferred for the manufacture of ferrites, especially yttrium iron garnet (YIG), each molecule of which contains 5 iron atoms (yttrium is one of the lesser known elements). Yttrium oxide is mixed with ferric oxide to produce YIG, after processing the material is pressed into the required shape and then fired as with pottery.

Looking further into electrons and their spins we find that if a d.c. magnetic field is applied to an electron, the latter precesses, in fact it acts like a gyroscope – see Figure 2.8(i). Precession is the slow movement of the axis of a spinning body but we leave it at that for study of the peculiarities of the gyroscope might lead to many sleepless nights. The electron precesses at a rotational frequency which is dependent on the strength of the applied magnetic field. Interaction between an incoming microwave and the ferrite electrons is maximum when the microwave frequency is at the precession frequency, known more technically as the *gyromatic resonance frequency*. We will not delve into this interaction, it is sufficient here to note that wave polarization (Sect.1.4) can be changed.

When a circularly polarized wave meets precessing electrons, it is found that there is interaction only when the wave rotation is in the same direction as the electron spin. Now it can be shown that linear polarization is made up of two equal circular polarizations rotating in opposite directions. Hence looking this way at a linearly polarized wave (e.g. TEM – Sect.1.4) when it is travelling through a ferrite in the same direction as the static magnetic field, the two circularly polarized waves representing it interact differently since one aids precession of the spinning electrons whereas the other opposes it. Accordingly their phase shifts are not equal and the polarization is rotated. This rather sketchy explanation is illustrated in a practical way in the next two sections.



(i) an electron precessing in a magnetic field

(ii) basic elements in a Faraday rotation isolator

(iii) action in the isolator

FIG. 2.8 POLARIZATION ROTATION BY A FERRITE

2.5.1 Isolators

It was Faraday himself who first saw the light (or did not as the case may be). He discovered that a linearly polarized light wave experienced rotation when transmitted through certain materials in a direction parallel to that of an applied magnetic flux. Microwaves as we noted in Section 1.6, are not far removed from light waves so we might reasonably expect the same effect with microwaves and this is so. A sketch of the basic elements of a *Faraday rotation isolator* is given in Figure 2.8(ii). In (iii) the rotation is further illustrated. When the electric field lines of the microwave are at right angles to the attenuation vanes there is little attenuation but this steadily increases as the angle decreases from 90° down to 0°. Accordingly we see from (iii) that because the attenuation vane of the second waveguide is placed at 45° to that of waveguide 1, a microwave travelling through 1 and 2 in that order suffers little attenuation. Conversely in the opposite direction the ferrite rotates the wave polarization so that the electric lines run parallel to the attenuation vane and the wave therefore experiences a high loss. Thus any reflections from the far-end (which we have seen from Section 2.1.2 create undesirable standing waves) are effectively removed.

There are many other configurations of isolators based on ferrites. One in common use is a resonance type and here the resonance is of the precessing frequency. Interaction is strong or negligible depending on the direction of propagation of the wave.

2.5.2 Circulators

The non-reciprocal phase-shifting characteristics of ferrites are again used to advantage in *circulators.* Perhaps the most frequently used component is the *3-port* or *Y* circulator as sketched in Figure 2.9. The device is effectively three isolators as described above built into one. Microwave power entering port 1 is received at 2 only, that entering at 2 is received at 3 only and that entering at 3 is received at 1. Typical uses are indicated on the diagram, i.e.:

(i) a transmitter at 1 feeds an antenna at 2. Power received from the antenna is fed to its receiver at 3;

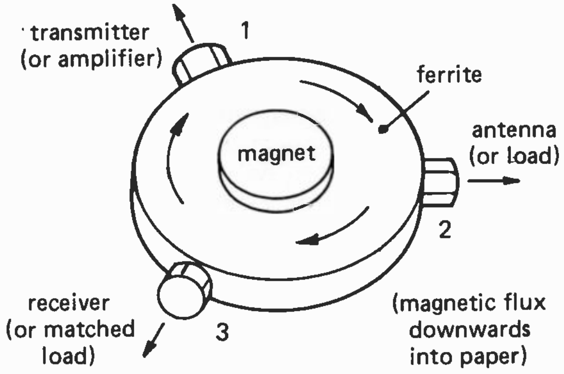


FIG. 2.9 3–PORT CIRCULATOR

(ii) an amplifier output connected to 1 is coupled to the load at 2. Any reflected power from a mismatched load re-enters the circulator at 2 and is dissipated in a load connected to 3.

Above are two examples only, circulators with more ports than 3 are in use and as microwave requirements grow, ferrite technology expands to satisfy them. Certainly ferrites have found their place in satellite television in which the electromagnetic wave transmitted down to us from the satellite may be either vertically or horizontally polarized (Sect.1.4) and the unwanted polarization must be rejected to prevent interference (see Sect.5.3). The *ferrite polarizer* is mounted on the receiving antenna (the "dish") and it can be remotely switched to allow only the required polarization to pass to the receiving equipment. Practically all antennas are mounted in the open so the fact that the device is robust and is itself waterproof is an indication of its superiority over many alternative systems.

2.5.3 Filters

Ferrites, especially of the yttrium iron garnet type are specially adaptable as band-stop and band-pass filters of narrow width. For this a single crystal of YIG is grown from a solution, it then goes through a tumbling process with a fine abrasive powder until a highly polished sphere is obtained. In use the r.f. magnetic field of the wave to be filtered is coupled to the crystal by a loop at right angles to the d.c. magnetic field which controls the gyromagnetic (precession) resonance frequency. At this frequency, because the YIG is a single crystal the energy of the r.f. wave becomes tightly coupled to the gyromagnetic resonance and is absorbed by it. This produces a band-stop filter which is tunable over a wide frequency range by adjustment of the d.c. magnetic field.

A band-pass filter using the same type of YIG crystal sphere is slightly more complicated in that there are two loops, one carrying the input r.f. wave, the second (output) close to it but at right angles. Because normally there is no coupling between the input and output loops, the device attenuates the r.f. signal. However, at the gyromagnetic resonance frequency (adjusted by the d.c. magnetic field) the two loops become coupled energy-wise within the crystal and the filter passes the wave but only over a narrow frequency band. An example of the use of a YIG band-pass filter is given in Section 4.1.

Chapter 3

MICROWAVE GENERATION AND PROCESSING

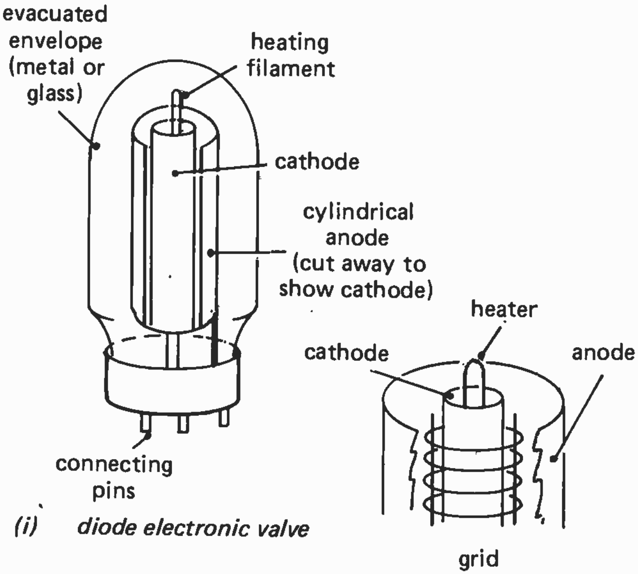
Although the transistor has rather indecently dismissed the vacuum tube (or electronic valve) in many electronics fields, it cannot do so yet for microwaves. Several different types of vacuum tube still hold sway when large powers are required, especially at the highest frequencies. At present solid state devices are mainly used at powers less than some 100 watts for frequencies up to 1 GHz, falling to a fraction of a watt at 100 GHz. On the other hand a vacuum tube device such as a klystron can handle as much as 1000 kW although only over a narrow frequency range. Such powers of course are seldom required.

We might conveniently divide the whole range of generators and amplifiers as in the sections below.

3.1 Grid-Controlled Microwave Tubes

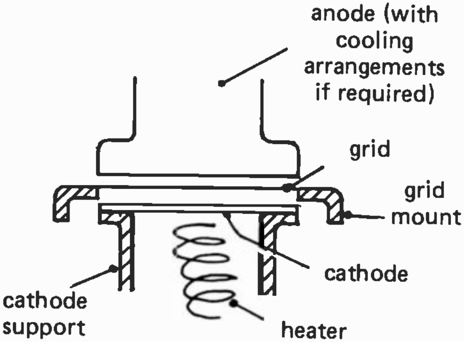
This type assumes a vacuum chamber through which a stream of electrons flows. All rely on thermionic emission of electrons from a *cathode*, i.e. the emission of electrons from the surface of a material when it is heated. At room temperature few free electrons have sufficient energy for escape from the surface. They are then prevented from moving away altogether by the space charge which they themselves create and by the fact that the atoms from which they have escaped now exhibit a positive charge. Heating the material provides more electrons with the energy they need for escape but accordingly the space charge increases. The restraining effect of the space charge is countered by surrounding the cathode by an *anode* which is charged positively. This positive charge attracts electrons from the space charge and provided that they have somewhere to go and can be replaced there is a flow of electrons from cathode to anode. Heat is provided by a heater element as sketched in Figure 3.1(i) which in fact illustrates a conventional diode electronic valve.

A fine-wire mesh *grid* inserted between cathode and anode as in (ii) enables us to control the electron stream, e.g. if a



(i) diode electronic valve

(ii) the addition of a grid forms a triode



(iii) electrode assembly of a planar triode

FIG. 3.1 GRID–CONTROLLED TUBES

negative potential exists on the grid, electrons are repelled but if the potential is positive electrons are encouraged to cross the gap. This arrangement constitutes the well-known triode valve which is basically of a tubular shape. When it comes to microwaves, however, the tubular triode has its problems for now the relationship between electrode dimensions and wavelength begins to show. The later developed *planar* triodes are better in this respect and the elements of such a device are shown in (iii) of the figure. They are especially useful in microwave generation where a crystal-controlled oscillator drives a multistage planar amplifier.

Summing up, the tubular and planar triodes function by space charge control in which the density of an electron current is varied. The performance tends to fall with frequency mainly because of the effects of electron transit time and interelectrode capacitances. Tubes which suffer less from these disadvantages function by modulating the *velocity* of the electron stream rather than its density as discussed next.

3.2 Linear Beam Microwave Tubes

These tubes are based on the manipulation of a high-velocity pencil-like beam of electrons. Each tube requires a generator of this type of beam, known as an *electron gun.* Such a device is not rare for it is from one of these that the picture on a television tube is ultimately derived. We might define an electron gun as an assembly of electrodes for the production of a concentrated beam of electrons moving at very high speed (many millions of miles or kilometres per hour) and therefore abounding with kinetic energy (k.e. = $\frac{1}{2}mv^2$ where m is the mass and v the velocity).

In the gun thermionic emission from a specially coated cathode provides the supply of free electrons, they are then accelerated towards the positive field of an anode as sketched in a simplified form in Figure 3.2. The cathode emitting surface is of considerably greater area than the cross-sectional area of the beam since the current density required in the beam is high. The focusing electrode is at cathode potential (or more negative) and serves to concentrate the beam as shown. The anode which is a metal plate with a circular hole in the centre is at several kilovolts positive relative to the

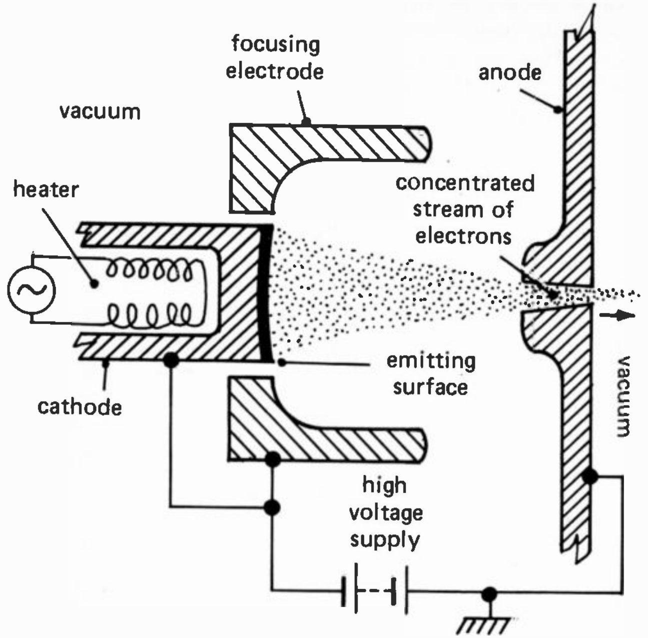


FIG. 3.2 SIMPLIFIED ELECTRON GUN FOR LINEAR BEAM TUBE

cathode. It is at earth potential so that output and input ports connected to the device supplied by the gun may also be at earth potential. Careful shaping of all the electrodes produces a convergence of the electrons so that they pass through the anode as a high-velocity narrow beam. The electron beam current (I) is given approximately by the relationship:

$$I = kV^{3/2}$$

where V is the anode-cathode voltage and k is a constant for the particular electrode structure.

Needless to say, practical guns are more complex than that shown in the figure and many variations are used depending on the particular device for which the gun is designed.

The beam is useless unless it can be modified or modulated in some way. This can be done based on the fact that each electron in the beam possesses kinetic energy because of its motion and if it enters an accelerating (positive) field it gains energy from the field and its velocity therefore increases. On the other hand, if it enters a retarding field the electron gives up energy to the field and its velocity decreases (here it is not essential for us to get involved with the minute corrections arising from Einstein and his relativity theory). Thus a radio frequency (r.f.) signal can be applied to an electron beam to alternately speed up and then slow down the passing electrons. This is appropriately known as *velocity modulation.* The practical outcome is the basis of operation of the *klystron, travelling-wave tube* and *backward-wave oscillator.*

3.2.1 Klystron

Klystrons come in all shapes and sizes, from the smallest some 10 cm long, to a large one of length about 2 metres. They can be used as microwave oscillators or perhaps more likely as amplifiers. The simplest klystron is a *two cavity* device (Sect.2.1.4), meaning that two resonant cavities are required, one to velocity modulate an electron beam, the other as the output system. These are known as *buncher* and *catcher* as shown in Figure 3.3. Between the two cavities is a length of tunnel known as the *drift space* or *drift tube.* Operation of the klystron follows from the electron gun and basic idea of velocity modulation given in the preceding section.

At one end of the evacuated tube is the electron gun, its beam passing first through the grids of the buncher cavity. The space between the grids is known as the *interaction space* and electrons passing through have their velocities changed according to the instantaneous r.f. potential existing across the cavity. As the electrons leave the interaction space therefore, some are travelling faster, some more slowly while others are unaffected. They are now traversing the drift space in which slowed electrons are overtaken by the faster ones hence creating bunches of electrons of density according to the magnitude of the input r.f. signal when the electrons were passing through the buncher.

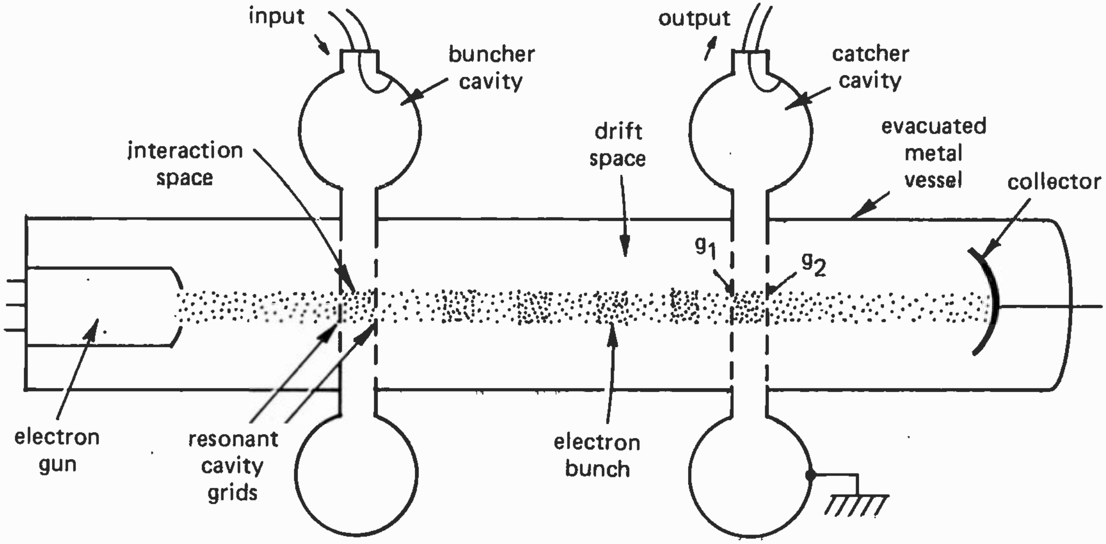


FIG. 3.3 TWO–CAVITY KLYSTRON AMPLIFIER

The catcher cavity is tuned to the same frequency as the buncher. It is located along the tube at the point where the electron bunching is maximum. Consider the two grids of the cavity, labelled g_1 and g_2 in the figure. If for example, due to an alternating field across the cavity g_1 is positive to g_2, then the bunches of electrons passing through will be slowed, hence as we have seen above, they deliver energy to the field which accordingly increases. In fact the phase relationship of the r.f. oscillation in the cavity relative to the arriving electron bunches adjusts so that the bunches are retarded because the oscillation amplitude cannot build up unless energy is extracted from the beam. The power output is related to the difference in average kinetic energy of the electrons as they pass from g_1 to g_2. Following the catcher cavity is the collector, simply a positive anode which removes the beam.

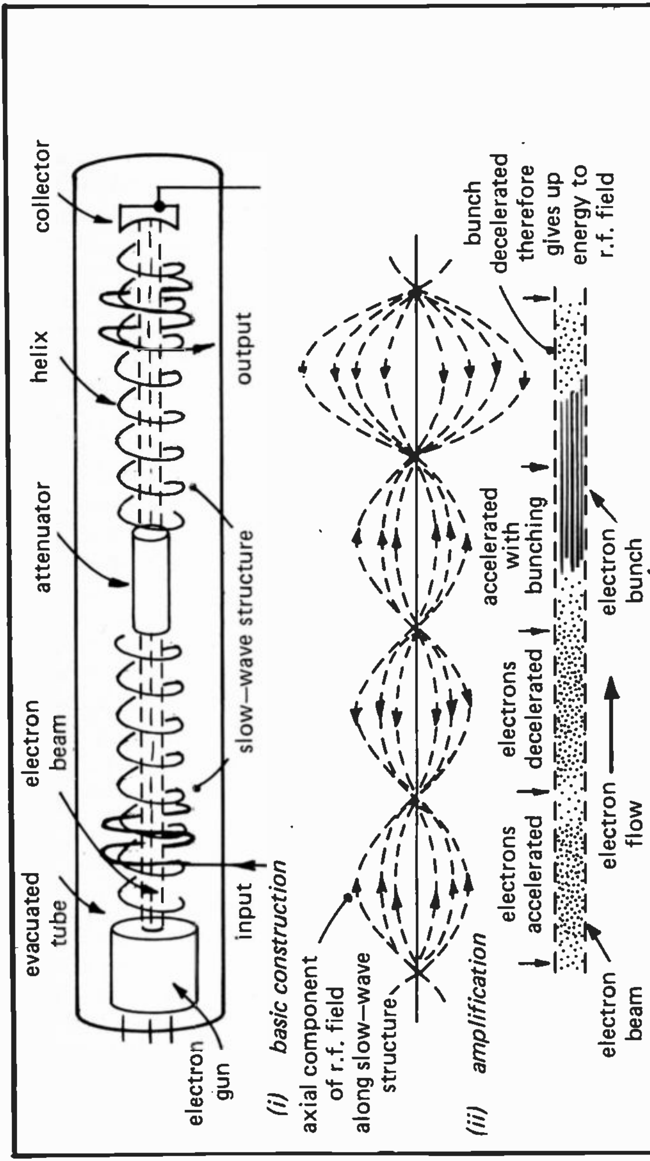
This is how a klystron amplifies: by arranging feedback from the output cavity to the input cavity in correct amplitude and phase the device becomes a microwave oscillator. Alternatively a special design of tube known as a *reflex klystron* may be used as a low power oscillator. This has only one cavity which serves as both buncher and catcher. The velocity modulated beam is returned to the cavity by a reflector electrode.

When high power is required, a klystron may have more than two resonant cavities, generally four but possibly more. Klystrons are sometimes used in telecommunications satellites but because as amplifiers they have a rather restricted bandwidth, travelling-wave tubes (see next section) are generally preferred.

3.2.2 Travelling-Wave Tube

This type differs from the klystron in that the r.f. input wave travels with the electron beam over almost its whole length and thereby is enabled to interact with it continuously instead of only in a short (buncher) cavity. As might be expected, therefore, as amplifiers they are capable of higher gain and because no resonant cavities are involved, considerably greater bandwidth is available.

Amplification by a travelling-wave tube (t.w.t.) arises from velocity modulation of an electron beam (Sect.3.2) by the r.f.


(i) basic construction

(ii) amplification

FIG. 3.4 TRAVELLING-WAVE TUBE

input wave which is followed by electron deceleration to pass energy back to the wave. A sketch of the basic arrangement is given in Figure 3.4(i). An electron gun (Sect.3.2 again) provides a narrow beam of electrons travelling lengthwise through the tube to a collector at the end remote from the gun. A metal helix surrounds the beam as shown. This helix carries the input r.f. and although the wave travels around the helix wire at a speed approaching that of light, its velocity along the tube axis is considerably slower depending on the pitch and angle of the helix. In fact the helix is known as the *slow-wave structure.* It is essential that the axial velocity is approximately equal to that of the electron beam and when this is so there is interaction between the beam and the r.f. wave.

The electron stream has a tendency to spread out due to the mutual repulsion between the electrons. To maintain them in a straight line an external longitudinal magnetic field is applied to repel the wanderers back into the stream; the field is usually provided by a permanent magnet but a coil may also be used (not shown in the figure).

Needless to say the action is quite complex but we can get some idea of how it all happens from Figure 3.4(ii), remembering that although the r.f. wave travels around the helix, it has an axial component which can react with the beam electrons. In the drawing this component and the electron beam are shown separated for clarity, the beam travels slightly faster. Electrons in the beam are alternately accelerated and decelerated, i.e. they are velocity modulated, therefore creating bunches. As a bunch progresses further along it is decelerated and so gives up energy to the field, resulting in an induced r.f. current in the helix. Provided that this is in the correct phase, the r.f. signal is increased and is now able to produce a denser electric charge in the beam (bunch). The effect is cumulative and the wave builds up until the average velocity of the electrons in the beam is reduced to that of the travelling field. As an amplifier the tube gain varies directly as the effective length of the slow-wave structure. Power gains of 60 dB are common.

We may be surprised to see of all things an attenuator in the centre of the slow-wave structure. It is a device which allows the electron beam with its bunches to travel through unopposed whereas the r.f. wave is reduced to almost zero. The technique prevents unwanted waves reflected from input mismatches from being propagated to the tube input and therefore causing instability. The amplified signal is removed from the helix by the output coupling. Note that the energy employed in amplifying the r.f. signal is provided entirely by the electron beam which can be looked upon as a d.c. source.

Alternative slow-wave structures may be employed or even a series of cavities, the basic method of operation however is the same. Travelling-wave tubes are extensively used as amplifiers in television broadcasting (both terrestrial and satellite), in microwave communication systems and in radar and military weapon systems.

3.2.3 Backward-Wave Oscillator

To have successfully coped with the operation of the travelling-wave tube (t.w.t.) may in fact make things a little difficult when trying to get to grips with the *backward-wave oscillator.* It seems to be a t.w.t. all over again yet the direction of energy flow in the slow-wave structure is in the opposite direction to that of the electrons in the beam. Figure 3.4(i) applies except that there is no attenuator, the electron beam has a hollow cross-section, the slow-wave structure consists of a flat metal tape and there is a single output coupling only at the electron gun end.

Visualizing the interaction process is not exactly child's play and the unadorned statement that the phase and group velocities of the r.f. circuit are 180° apart may be of little help. Nevertheless, from this it follows that the direction of energy flow (the group velocity) is directly opposite to the direction of electron motion. This shows that a wave is propagated backwards. We can look a little more closely into the process by considering a section of the slow-wave structure as in Figure 3.5. As with other oscillators, noise voltages (e.g. from the electron beam) induced in the slow-wave structure are sufficient to start a build-up of oscillations.

In the figure the electric field existing across adjacent helix turns is shown. As mentioned in Section 2.3.1, an electric field cannot exist parallel and close to a conducting surface, accordingly under each turn of the helix the field is zero with

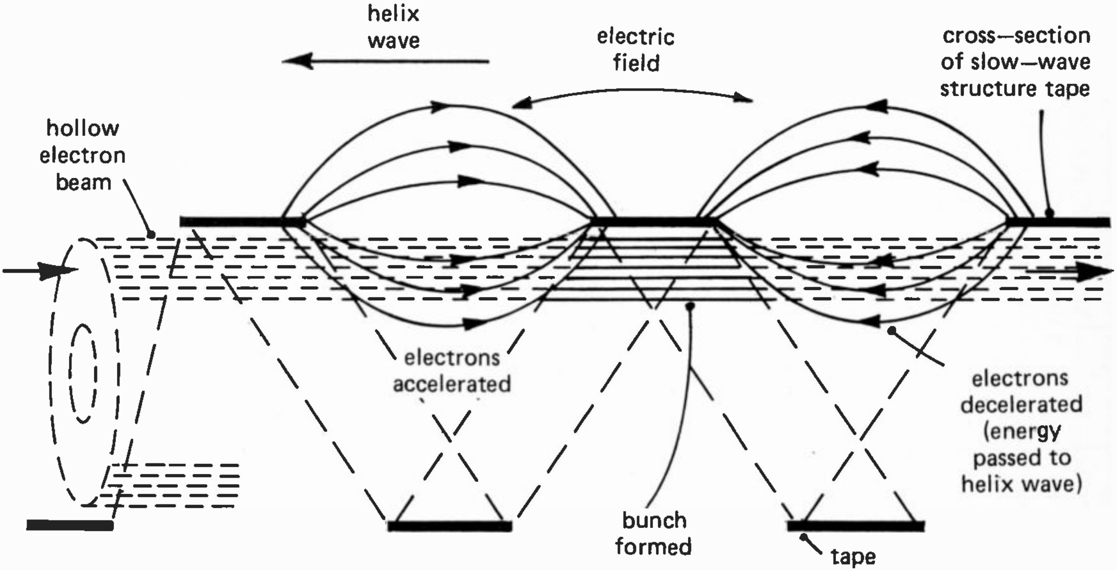


FIG. 3.5 INTERACTION BETWEEN HELIX WAVE AND ELECTRON BEAM

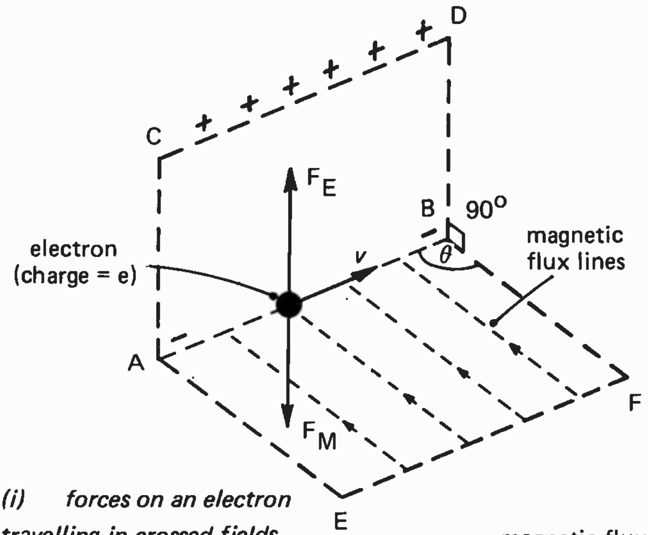
a maximum midway between the turns as shown. As with the t.w.t. the electrons in the beam are alternately accelerated and decelerated. In a way this transfer of energy can be looked upon as a feedback system between mid-points of adjacent helix gaps and provided that phase relationships are correctly maintained by adjustment of beam velocity, the generated wave flows backwards and the system can be made to oscillate. It is also necessary to ensure that each group of electrons is affected by an electric field of the same phase as it passes along the slow-wave structure so that one group is continually accelerated while the following group is decelerated. The backward helix wave builds up and flows in a direction opposite to that of the electron beam and is extracted through a coupling at the gun end of the tube.

Low-power backward-wave oscillators are those having power outputs ranging from 10 to about 100 mW with operating frequencies extending over the whole of the microwave band. Power backward-wave oscillators are more likely to employ mutually coupled cavities instead of a metal tape helix. Output powers range from about 1 mW to 40 W over a frequency range from some 20 to 300 GHz.

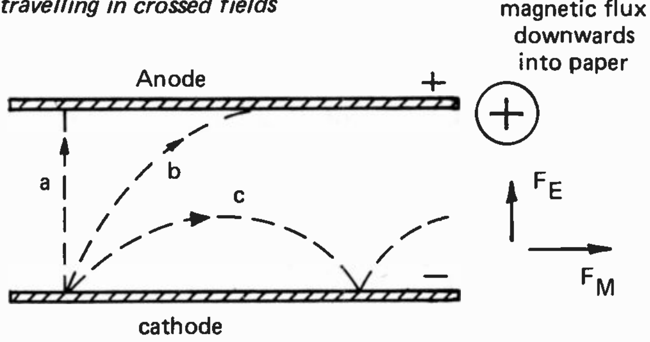
3.3 Crossed-Field Microwave Devices

As its name suggests, a linear beam tube is based on an undeviating stream of electrons. A magnetic field is involved but this is merely to keep the beam in order. The crossed-field type also employs a magnetic field but its purpose is now more fundamental, not to keep the electrons from straying from a straight and narrow path but in fact to encourage them to do so. Perhaps we should first refresh our memories on what a field can do to an electron.

In Figure 3.6(i) suppose an electron of charge e is travelling along the line from A to B at a velocity v through the two fields at right angles, one electric, the other magnetic (the crossed fields) as illustrated. The magnetic field is horizontal, has a flux density B and is in the direction shown by the arrows on the flux lines. The electric field is vertical and the flux lines have been omitted because shown conventionally they would be arrowed downwards which may be confusing.



(i) forces on an electron travelling in crossed fields



(ii) some electron paths

FIG. 3.6 ELECTRONS IN CROSSED FIELDS

It is sufficient to postulate that along the line CD is a positive charge. Using a right-hand rule (i.e. with fingers extended as originally suggested by Fleming – this is electron flow, not conventional current), there is a force F_M acting downwards on the electron due to the magnetic field such that with $\theta = 90°$:

$$F_M = evB$$

and we note that the force varies with the velocity of the electron. If θ is other than 90° then the formula for F_M becomes $evB \sin \theta$. In the opposite direction the electric field exerts a force F_E upwards on the electron such that $F_E = eE$ where E is the electric field strength.

If we wish our electron to continue its journey undisturbed then:

$$eE = evB$$

hence:

$$v = E/B.$$

When this condition does not apply the electron is deflected from its straight path. From this we now may begin to see the basis on which a crossed-field device works. The electric field is at microwave frequency and the magnetic field is provided (usually) by a permanent magnet. The combination of the electric and magnetic fields is used to make electrons deviate in many curvaceous ways. Here is just one simple example, illustrated by Figure 3.6(ii). In this case we must imagine the magnetic flux as being downwards into the paper. The line CD in (i) of the figure is now an anode and AB is a cathode. With zero magnetic field the electrons follow the path (a). As the magnetic field is increased, electrons move in a curved path (b). A further increase in the field prevents electrons from reaching the anode and they return to the cathode. We will see that many other gyrations are possible which may seem pointless until we recall that accelerating electrons take energy on board whereas when decelerating, they give it up.

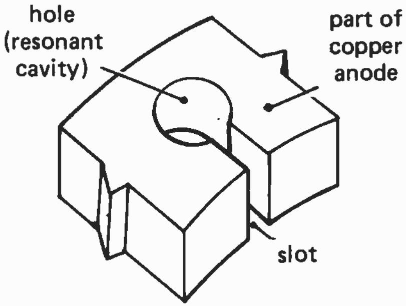
3.3.1 Magnetron

This is perhaps the best known of all crossed-field devices, mainly because of its early development for use in radar systems. It has now entered the home as the power behind (or in) the microwave cooker. Many types of magnetron exist, one which is in common use and lends itself to (moderately) simple explanation is the 8-cavity magnetron oscillator as illustrated in part in Figure 3.7. The anode is a circular block of metal (usually copper) in which are cut 8 holes and slots as shown in (i). The holes are in fact cavities all resonating at the same frequency and to one is coupled a waveguide as the output port. The cathode is fixed in the centre of the block, leaving an interaction cavity between it and the inner surface of the anode as shown in (ii). A permanent magnet is usually employed to create the magnetic field, the "lines" of which are downwards into the paper in this particular drawing.

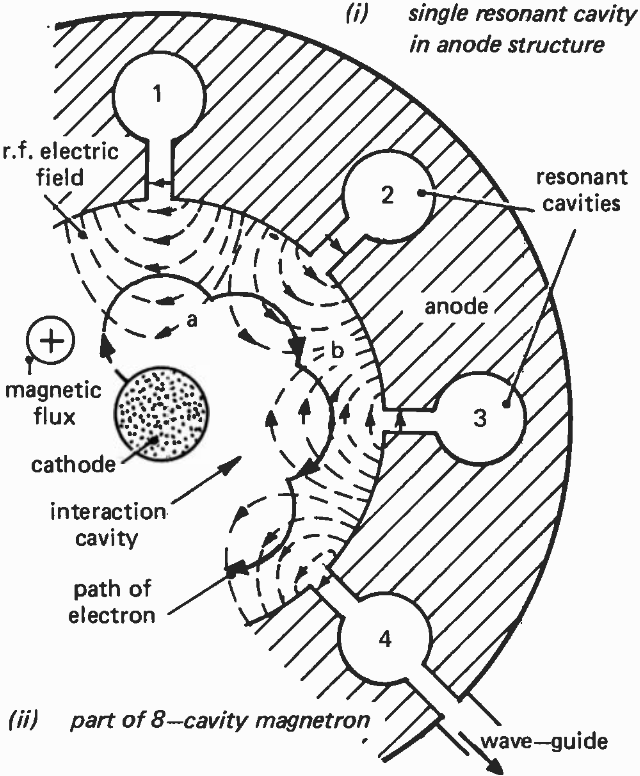
Some random effect sets up weak oscillations in the cavity resonators and these give rise to electric fields crossing the slots and extending into the interaction space as shown in the sketch *for a particular instant.* In this arrangement alternate slots support fields in opposite directions, the magnetron is then said to operate in the π-mode.

Magnetron operation is highly complex so the simplified explanation which follows can only serve to furnish an appreciation of the ingenuity on which it is founded. We cannot expect to understand it in detail; even the mechanical and electronic methods adopted for tuning the cavity structure would require many chapters on their own.

Let us consider the path of an electron liberated from the cathode. From Section 3.3 it clearly does not move in a straight line directly to the positive anode because the (d.c.) magnetic flux causes it to curve to the right (the right-hand rule again). Our electron enters the r.f. electric field of cavity 1 which is in the direction shown. The direction is such as to retard the electron which therefore transfers energy to the field. The electron slows towards point *a* but is again accelerated by the d.c. field. Now it comes under the influence of the r.f. field of cavity 2 but note that while the electron has travelled to *a*, the r.f. field has changed by 180° (π radians)



(i) single resonant cavity in anode structure



(ii) part of 8–cavity magnetron

FIG. 3.7 MAGNETRON

and is therefore in the opposite direction to that shown in (ii) of the figure. The process which we observed under cavity 1 is now repeated under cavity 2 and electron energy is again transferred to the field as the electron moves on to *b*. At cavity 3 the r.f. field has passed through 360° so is as shown and the same interaction continues. As energy is given up, gradually the electron moves closer to the anode and is finally seized by it, although this may be after many orbits around the cathode. Other electrons may not be so useful as agents in the energy transfer process. Some are soon collected by the anode, some revert to the cathode. Those which find themselves in an unfavourable position for energy provision because they are in the alternate phase of the r.f. field absorb energy are therefore accelerated and move back towards the cathode. Nevertheless, the whole process can be considered as a cloud of electrons rotating round the cathode in synchronism with the r.f. field and delivering energy to it.

Commercial magnetrons are available with mean output powers of between 10 and 100 kW at 1 GHz but down to a fraction of 1 kW at 100 GHz. Peak output powers however may be several hundred times greater.

3.3.2 Crossed-Field Amplifiers

Of the many varieties of crossed-field amplifier which exist, one which is commonly used is in fact quite similar to the magnetron. Actually the crossed-field amplifier is a logical extension of the magnetron (an oscillator) into an amplifier. As an amplifier there must of course be both input and output ports. Figure 3.8 is an imaginary crossed-field amplifier drawn to show the main basic operating features, practical devices may look very different.

Conversion of an 8-cavity magnetron into a crossed-field amplifier requires special conducting straps connected as shown to reduce the build-up of unwanted modes (Sect.2.3.2). The straps are not simply electrical connections (the anode block is usually solid copper anyway) but are recessed into the anode structure as shown in the cross-sectional view. For a π-mode device as described in the preceding section, the straps join segments which are at the same r.f. voltage as shown. Each strap has inductance and also forms a capacitance

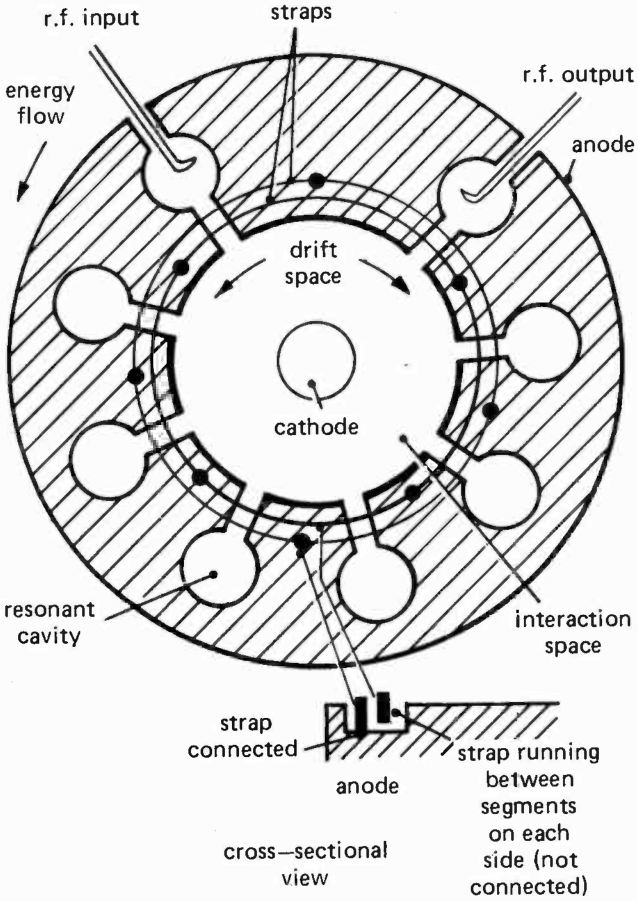


FIG. 3.8 CROSSED–FIELD AMPLIFIER

coupled to the segment in which it is recessed. The capacitance is effectively in parallel with the resonator hence reduces slightly the resonance frequency of the π-mode. However, for other lower modes it can be shown that there is an increase in the resonance frequency and so the separation between the wanted and unwanted modes is increased.

Note the increased interaction space (the *drift space*) between the input and output ports. Its purpose is to minimize feedback from output to input which would cause the device to fail as an amplifier because of oscillation. In passing through this space electron bunches in the rotating cloud of electrons travelling past the output port have time to disperse before they reach the input port.

Peak output power levels of the various crossed-field amplifiers available are surprisingly high. They vary over a wide range from between 100 kW and 5 MW at 1 GHz to 100 kW at 16 GHz. Gain of a single unit is of the order of 15 dB.

3.4 Semiconductor Devices

Naturally the development of semiconductor devices for use at microwave frequencies has made rapid strides. The main difficulty is perhaps obvious, semiconductor junctions have capacitance which at the higher frequencies is ruinous. Accordingly the earliest attempts to enlist semiconductor technology revolved around the point-contact diode which uses a metal *whisker* in contact with a semiconducting material and therefore has comparatively low junction capacitance. More recently special semiconductor components have been developed for microwaves. This cannot be an exhaustive survey of the whole field but a few examples may be useful here to demonstrate the underlying techniques.

Transistors we recall are broadly divided into two types: (i) *bipolar* which depend for their action on the flow of both majority and minority carriers with their terminals labelled emitter, base and collector; and (ii) *field-effect*, also known as *unipolar*, and which depend on the flow of majority carriers only. Their terminals are labelled source, gate and drain. We will see that much development for microwave use has centred around special constructions of the field-effect transistor (f.e.t.).

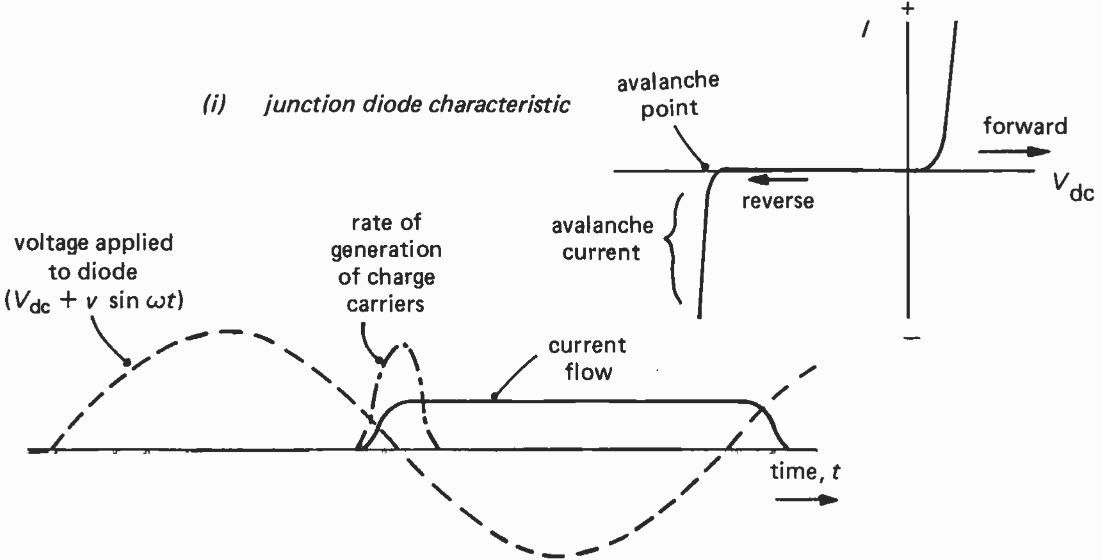
3.4.1 Diodes

Certain diodes are capable of amplification or sustaining oscillation if they have a negative resistance characteristic over the frequency range required. This needs some explanation so first we refresh our memories with the basic p-n junction current/voltage characteristic extending into the avalanche breakdown region as shown in Figure 3.9(i). When the d.c. voltage (V_{dc}) applied exceeds the breakdown voltage, avalanche multiplication generates an additional supply of electrons and holes and these flow in the depletion region which increases in width on the n-side of the junction according to the magnitude of the voltage applied.

Consider a diode maintained in the avalanche condition by a voltage V_{dc} and with an r.f. voltage $v \sin \omega t$ acting as shown in (ii) of the figure. The supply of charge carriers generated will vary as the net voltage applied but not in phase with it because a negative resistance condition exists where the current decreases as the voltage increases. In effect the maximum rate of supply of charge carriers occurs at over 90° *after* the applied voltage maximum. Following this the electrons drift across the depletion region and while doing so they constitute a current in the external resonant circuit. Thus the total delay of avalanche plus drift time ensures that the current is flowing and therefore energy is being delivered to the external circuit even while the applied r.f. voltage is in its negative half-cycle – see (ii) of the figure (don't be misled by the standard use of the term "drift", the total delay time at 10 GHz might be a mere 40 picoseconds). Most of the energy is provided by the circuit power supply.

Three diodes currently in use as microwave amplifiers and/or oscillators which are worthy of mention here are:

1. *tunnel diode:* this type is constructed with heavy doping on both sides of the junction resulting in a very narrow potential barrier. Accordingly charge carriers can "tunnel" across the depletion layer even when no potential difference is applied. Put in other words, electrons which normally



(i) junction diode characteristic

(ii) events in a cycle in a diode under avalanche breakdown

FIG. 3.9 AVALANCHE DIODES

would not have sufficient kinetic energy for overcoming the potential barrier, now have no need of it, the energies they possess as valence electrons are sufficient for them to leave home and jump the barrier as conduction electrons. With forward bias a negative resistance region develops which is voltage controlled. Tunnel diodes are sufficiently fast acting to be useful for microwave amplification (but at very low power) and for microwave oscillators.

2. *IMPATT diode:* the acronym IMPATT is derived from *IMP*act ionization *A*valanche *T*ransit *T*ime which is intended to define how the device works. Its main use is in the generation of microwave power and operation is based mainly on the principles outlined above, i.e. as a diode reverse-biased into avalanche breakdown; typically V_{dc} is around 100 V. In use the diode is mounted in a resonant cavity with impedance matching. Typically an IMPATT amplifier might have a maximum power output of about 100 W at 10 GHz with a gain of 10 – 12 dB. Efficiencies are fairly low at 10 – 20%. As an oscillator output powers approaching 100 W are possible at up to 10 GHz but generally lower than this at higher frequencies.

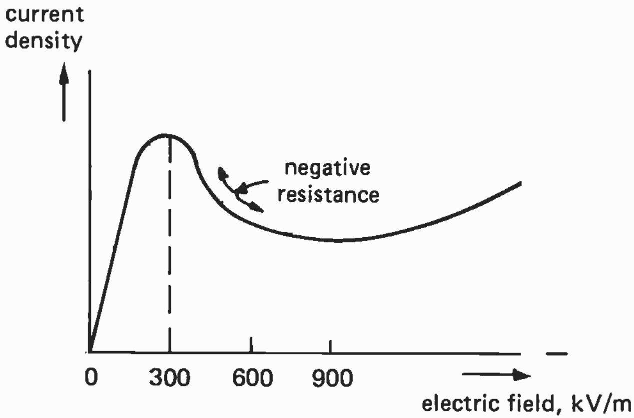
3. *Transferred Electron (Gunn) Device:* in this the special effect arises in a single semiconductor material (i.e. there is no p-n junction) usually of gallium arsenide or indium phosphide. The basic mechanism is different from that in "normal" diodes and it was first reported by J. B. Gunn in 1963 hence it is also known as a Gunn device or Gunn diode.

Gunn discovered that a d.c. voltage applied to a bar of n-type gallium arsenide or indium phosphide at first produced a current increasing linearly with the voltage but above a certain threshold voltage the current oscillated. The threshold voltage is around 300 kV/m which seems exceedingly high until we realize that practical devices may have an effective length as small as 0.1 mm, hence requiring a voltage of about 30 upwards. Gunn found that the current oscillations were closely related to the transit time of the electrons along the slice. As an example, for gallium arsenide the maximum electron drift velocity (v) is about 10^5 m/s so for a 0.1 mm length (l) the oscillation period (t) $\simeq l/v = 10^{-9}$ seconds and the frequency $= 1/t$, i.e. about 1 GHz. The necessary negative resistance feature is shown by the current density/electric field characteristic shown in Figure 3.10(i) which we can interpret for a given specimen as a current/voltage characteristic, the negative resistance is evident above about $V = 300$ kV/m.

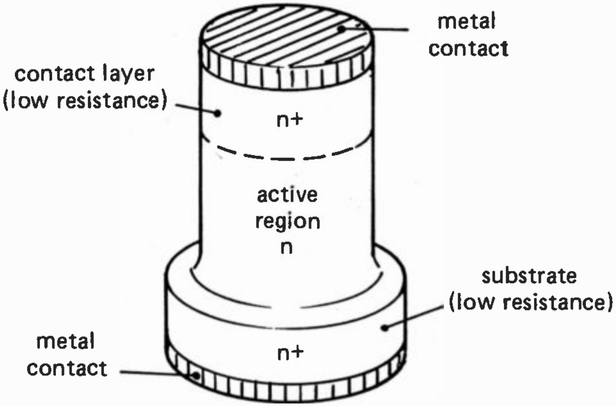
The basic mechanism is complex and is usually explained in terms of electron energy levels but it may be possible to visualize it in a rather simplified fashion through consideration of the space charge (Sect.3.1). Picture a small slice of gallium arsenide connected to a d.c. voltage supply which exceeds the threshold voltage. At the negative contact (the cathode) a space charge builds up with a consequent increase in its electric field. As it does so the voltage across it increases hence the voltage across the remaining region towards the positive contact (the anode) falls below the threshold value. The accumulation of space charge is known as a *mature domain*. Now this is taking place as part of a moving electron stream and in a way we can imagine the domain as being swept along with the current increasing and then decreasing and finally flowing out at the anode, hence the name *transferred electron device.* The process continually repeats at microwave frequency.

In the practical device the active region as shown in (ii) of the figure is an n-layer whereas the "contacts" are n-regions, highly doped for good conductivity (marked n^+). Oscillators of this type may have continuous wave outputs of up to 1 W at 10 GHz falling to around 100 mW at 40 GHz. Noise generated is low compared with most other types of microwave oscillators. Practical amplifiers may have gains of 10 – 30 dB over the range 10 – 50 GHz, depending on the number of stages.

Demodulation or *detection* is the process through which information impressed on a carrier wave is regained. For amplitude modulation systems the basis of most detectors is the non-linear characteristic of a diode. For frequency modulation systems (e.g. microwave telephony and satellite) the circuit is known as a *frequency discriminator* and is more



(i) current density/electric field characteristic for gallium arsenide



(ii) typical construction

FIG. 3.10 TRANSFERRED ELECTRON DEVICE

complicated but is still likely to be based partly on diodes. Two suitable types of diode are mentioned below:

4. *Schottky diodes:* microwave p-n diodes are limited at the higher frequencies because of the relatively long lifetimes of the minority carriers. *Lifetime* is the mean time interval between generation and recombination of the charge carriers and in the diode this means that it will not respond sufficiently quickly hence in a detector resulting in reduced efficiency. The Schottky diode relies for its rectifying action on a barrier formed at the junction between a metal and a semiconductor, e.g. aluminium or nickel-chromium with n-type silicon. With such a junction the current flow is mainly by majority carriers, hence minority carrier diffusion is small and less of an embarrassment at the higher microwave frequencies. Special mesh constructions allow operation up to about 40 GHz, with gallium arsenide it is possible to exceed 100 GHz.

5. *Backward diodes* are similar to tunnel diodes (see 1) except that the doping levels are slightly lower. The current/voltage characteristic is such that at low reverse bias (almost zero) the diode breaks down. In the forward direction at low voltages the characteristic is very non-linear and it is this which makes the diode suitable for use as a detector. Being a majority carrier device the diode has a high speed of response and is usable up to at least 20 GHz.

3.4.2 Bipolar Transistors

Compared with the semiconductor diode the additional electrode gives more control. Arranging for microwave purposes mainly involves keeping reactances (mostly capacitive) and parasitic resistances as low as possible. Accordingly microwave bipolar transistors are manufactured using planar technology and are of extremely small dimensions, e.g. with active surfaces of less than one thousandth of a square millimetre. Needless to say, except for power transistors the power handling capacity of such a tiny device is small. One advantage of the microwave bipolar transistor is its low cost. Operation of silicon types extends up to some 6 GHz for both amplifiers and oscillators.

3.4.3 Field-Effect Transistors

These as normally used have little more to offer than the bipolar. However, specially constructed field-effect transistors (f.e.t.'s) are capable of operation up to 100 GHz although useful amplifier gains are limited to around 10 GHz. The special construction results in the device being known as a MESFET which is an acronym derived from metal-semiconductor f.e.t. The simplicity of a MESFET can be judged from the diagrammatic sketch of Figure 3.11. The construction is similar to that of a conventional f.e.t except that the p-n junction at the gate electrode is changed to a metal-semiconductor junction, in this case aluminium (Al) to n-type gallium arsenide (GaAs).

We need not delve too deeply into the significance of *Fermi levels*, it is sufficient here to say that this is the particular level of energy for a material at or just below which electrons contribute most to the conductivity. The Fermi level of gallium arsenide is greater than that of aluminium, accordingly there are more available free electrons in the GaAs. At the area of contact the Fermi levels must be equal hence the higher energy electrons of the GaAs flow into the aluminium so leaving a positive charge in the semiconductor and a negative charge in the metal at the junction. A junction barrier potential equal to the initial Fermi level difference arises and a depletion region is formed in the semiconductor owing to the loss of charge carriers.

If now a potential is applied between source and gate the charge pattern is modified. The depth of the depletion region must also change since its volume depends on the junction barrier potential. Majority carrier flow (electrons) is from source to drain and since the depletion region restricts the channel available for electron flow the drain current is similarly modified. Thus r.f. potentials applied to the gate appear amplified in the load. Note that since the d.c. potential along the bar rises positively from source to drain, the depletion layer is not of constant depth but varies as suggested in the

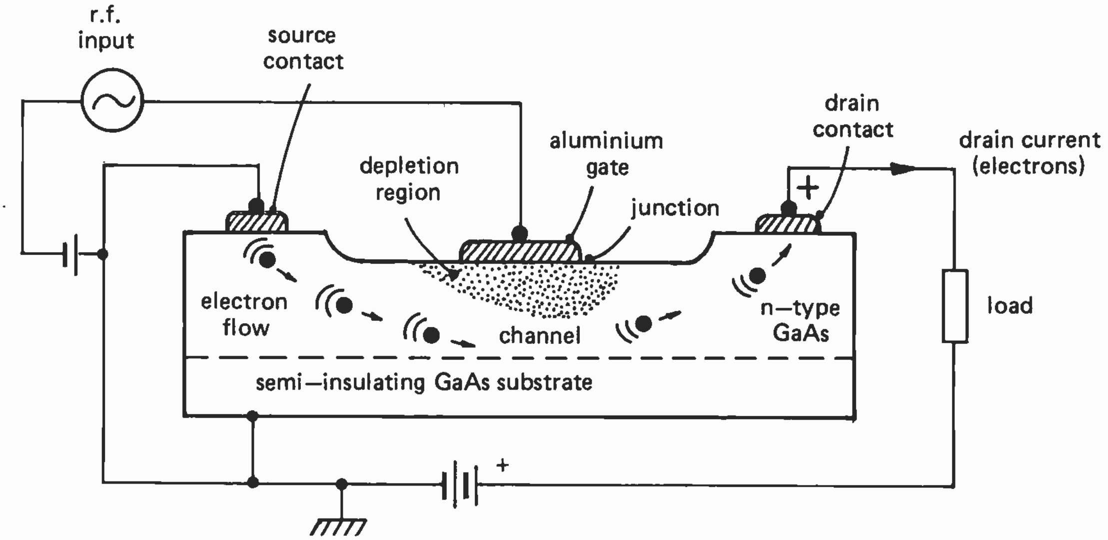


FIG. 3.11 CONSTRUCTION OF A MESFET

drawing.

These devices can be made very small, hence electron transit time is short. In addition, electron mobility in GaAs is 3 to 4 times that in silicon. Electron transit time and electron mobility are key factors in determining the high frequency response of a semiconductor device (or in fact, any device). When a high frequency wave initiates action (usually a flow of electrons) within a device, this must be completed as far as possible before the wave polarity reverses. Electron velocity through a semiconductor device increases with electron mobility as might be expected. High velocities (some 10^5 m/s for GaAs) result in shorter transit times. GaAs is superior to silicon in this respect hence the GaAs MESFET has better performance at microwave frequencies compared with the normal silicon f.e.t.

These MESFET's being basically transistors are usable in a variety of applications, especially in low-noise integrated circuit (IC) amplifiers and oscillators. For IC's the general classification MMIC (monolithic microwave integrated circuit) may be used. Power MESFET's can be connected in Class B for higher efficiencies. Their low cost and good performance make them replacements for other devices such as diodes and even the travelling-wave tube.

Research into GaAs microwave integrated circuits (GaAs MMIC's) continues apace and complete circuits on a chip are available working at frequencies in excess of 20 GHz. The packing density can be very high, for example a 6 GHz chip containing microwave filters, amplifiers, output amplifier and routing switches (some 40 f.e.t.'s + 150 components) can be realized on a chip of area less than 10 mm^2. Before MMIC's came along "breadboard" designing allowed the insertion of adjustable tuners, capacitors, potentiometers, etc., so that circuit action could be varied to maximize performance. Naturally MMIC's are not likely to permit such arrangements, hence their design is more complex and computer aid is usually required. In addition, it perhaps goes without saying that owing to the extremely small size of microwave devices, circuit analysis and testing present many difficulties.

3.4.4 Parametric Amplifiers

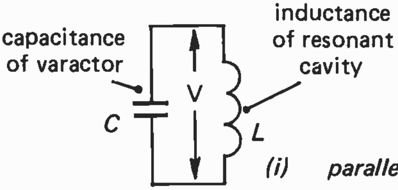
"Parametric" implies that at least one parameter in the system is non-linear or varies with time. On this basis microwave amplifiers are designed around the variation in capacitance of a *varactor* diode. This small, specially constructed semiconductor diode is located in a suitable microwave resonant cavity and the energy source for amplification is derived from a high frequency wave (the *pump* frequency) instead of from a d.c. supply as with most other systems. We revise our understanding of the varactor diode first. The varactor diode (or just plain *varactor* – from variable reactor) is well known for its use as an electronically controlled tuning capacitor in a radio or television receiver. Its capacitance is small but this is not an embarrassment in microwave applications.

In any reverse-biased semiconductor diode the depletion layer contains very few charge carriers so the layer has a high resistance (e.g. > 10 MΩ) and is effectively a dielectric between two conducting plates, the n- and p-type regions. The varactor therefore acts as a parallel plate capacitor. What is important is that the width of the depletion layer varies with the applied reverse voltage, hence the device can be used as a voltage controlled capacitor, the capacitance falling as the reverse voltage is increased. At microwave frequencies only a picofarad or so may be required with a variation perhaps in excess of half a picofarad for a 4 – 6 V change; there is a wide variation however, so much depends on the method of fabrication. Certainly gallium arsenide is preferred as the semiconductor for it has a higher frequency response compared with silicon because of its higher electron mobility. It also has a lower series resistance (i.e. the resistance through the semiconducting materials to the depletion layer) which increases the efficiency since very approximately:

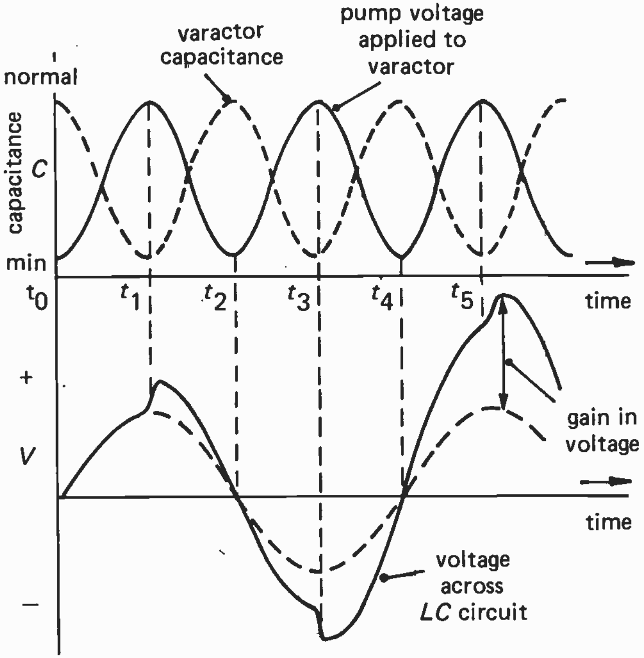
$$Q = \frac{1}{2\pi f r_s C}$$

where f is the frequency of operation, r_s is the series resistance and C the capacitance.

The parametric amplifier was developed in the USA and it cunningly exploits the fact that with any variable capacitor



(i) parallel LC circuit



(ii) showing how V increases when the varactor capacitance falls

FIG. 3.12 PARAMETRIC AMPLIFICATION

(such as a varactor), if at any instant the capacitance changes, then the voltage (V) across it must also change since nothing has happened to the charge (q) and $V = q/C$.

The explanation which follows is simplified so that the basic technology can be revealed. Practical parametric amplifiers need several additional components to make it all happen. Consider a parallel tuned circuit as representing the capacitance of a varactor diode within a microwave resonant cavity as shown in Figure 3.12(i). A frequency at twice that of the output is applied to the varactor, this is the pump frequency. Suppose the varactor capacitance is "normal" at times t_0, t_2, t_4, etc., but falls to a certain minimum at t_1, t_3, t_5, etc. From t_0 to t_1 the voltage, V across the LC circuit rises but note that by t_1 the varactor capacitance has decreased, hence from $V = q/C$, V rises. The signal frequency therefore experiences a voltage gain, the energy for this being supplied by the pump frequency. Note that the upper pair of curves in (ii) of the figure assume a linear relationship between reverse voltage applied to a varactor and the junction capacitance. This is certainly not the case but the assumption is made for clarity.

At t_2 the pump voltage is minimum so that the capacitance restores to its normal value and this is exactly when the signal voltage passes through zero, hence it is unaffected. At t_3 the pump voltage is maximum with the capacitance falling, hence V gains another increase as shown. There is no change at t_4 but this is followed by another increase at t_5. It is evident that the pump voltage must be applied in the correct phase otherwise such gains will not be realized and the LC circuit oscillation will be damped.

Practical circuits involve more stages than the above explanation would seem to indicate, especially in that the input and amplified output signals must be kept separate and for this a circulator (Sect.2.5) is employed. For basic amplifiers we have seen that the capacitance changes must occur at exactly the right times so to ensure this a second resonant (*idler*) circuit may be employed. However, in spite of what has been said, there are other systems in which the pump frequency differs from twice the signal frequency.

There is a maximum frequency of operation, it is greatly affected by the varactor series resistance (r_s). The formula is simply:

$$f_c = \frac{1}{2\pi r_s C_b}$$

where f_c is the frequency of cut-off and C_b the capacitance at breakdown. We compare two practical diodes, both having $C_b = 0.11$ pF but A has $r_s = 48\ \Omega$ whereas for B, $r_s = 5.3\ \Omega$. Then for A, $f_c = 30$ GHz but for B, $f_c = 273$ GHz showing that for a high signal frequency the varactor must be constructed with r_s as low as possible.

The parametric amplifier is noted for its low noise performance, it is even better if the device is cooled. A 20 dB gain amplifier might have a bandwidth of around 2% of the operating frequency, wider bandwidths can be achieved with lower gains.

3.5 Mixers

Communication by microwave radio invariably employs some sort of *frequency-changing* or *conversion*, e.g. a complex wave carried by a frequency f_1 may be centred on a new frequency f_2 with none of the information contained in f_1 changed in any way. Reasonably the device or circuit employed is known as a *frequency changer* but because it involves the mixing of two frequencies it is perhaps more commonly known as a *mixer*. This term is not fully descriptive of the action for merely adding two waves together does not produce the effect required unless they are mixed non-linearly. To be more precise, mixing is defined as the conversion of a signal of low power from one frequency to another by combining it with a signal at higher power in a non-linear device. The higher powered signal (f_0) is generated by a *local oscillator*.

Mixing produces a whole host of sum and difference frequencies but the one usually picked out is the straightforward difference between the two being mixed, i.e.:

$$f_2 = f_0 - f_1 \cdot$$

Microwave semiconductor diodes (Sect.3.4.1) are well suited for microwave mixing and a rectifier bridge circuit in common use is that of the *double-balanced mixer* which well suppresses in the output both f_0 and f_1 and certain harmonics, thereby reducing the complexity of subsequent filtering. The main elements of such a circuit are shown in Figure 3.13.

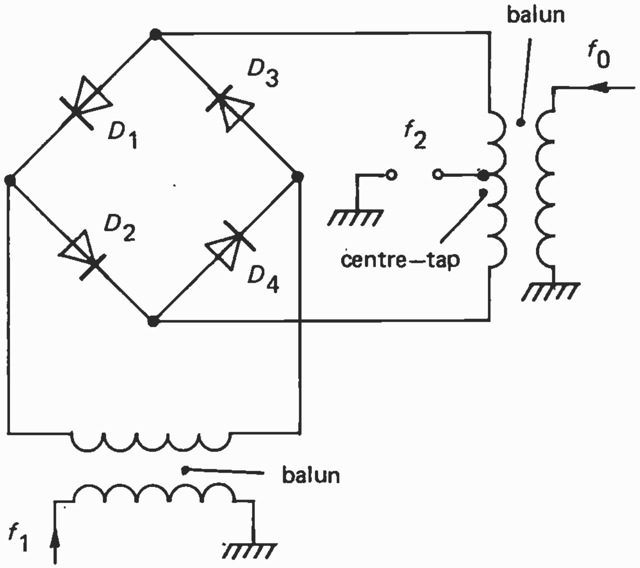


FIG. 3.13 DOUBLE–BALANCED MIXER

Baluns are employed at the inputs of f_0 and f_1. The term is an acronym from *balanced-to-unbalanced* (or vice versa), indicating a type of transformer used to couple a balanced impedance (as is the double-balanced mixer), to an unbalanced input such as a coaxial cable. As an example, a simple balun may consist of a short-circuited quarter-wave section of a transmission line attached to the outer braiding of the cable. This is arranged to present a very high impedance (Sect.2.1.3) to any wave arriving via the cable outer conductor, in other words, effectively disconnecting its coupling to earth.

Very briefly, it can be seen from Figure 3.13 that the large voltage of f_0 alternately switches D_1 and D_2, then D_3 and D_4 "on" and "off" thereby "chopping" f_1 at the frequency f_0. The circuit also shows clearly that neither f_0 nor f_1 is transmitted to the output terminals (f_2).

Mixing is part of the superheterodyne principle which is the backbone of radio receivers generally, f_1 is the incoming signal frequency and the mixer output (which we have designated f_2) is known as the *intermediate frequency* (see also Sect.4.1). Although the above explanation is centred on suitable microwave diodes, other types of mixers are worthy of mention. Certain transistors can be used with mixing taking place across the base-emitter junction so that f_2 is subsequently amplified by transistor action. Compared with diodes however, the transistor is likely to have an inferior noise performance. Of interest too is the *parametric converter*, similar to the parametric amplifier (Sect.3.4.4) but adapted for frequency conversion. Varactor diodes are semiconductor diodes which behave as voltage-dependent capacitors and these may be employed as with the amplifier with their capacitances varied on a time basis by the local oscillator (the pump).

Chapter 4

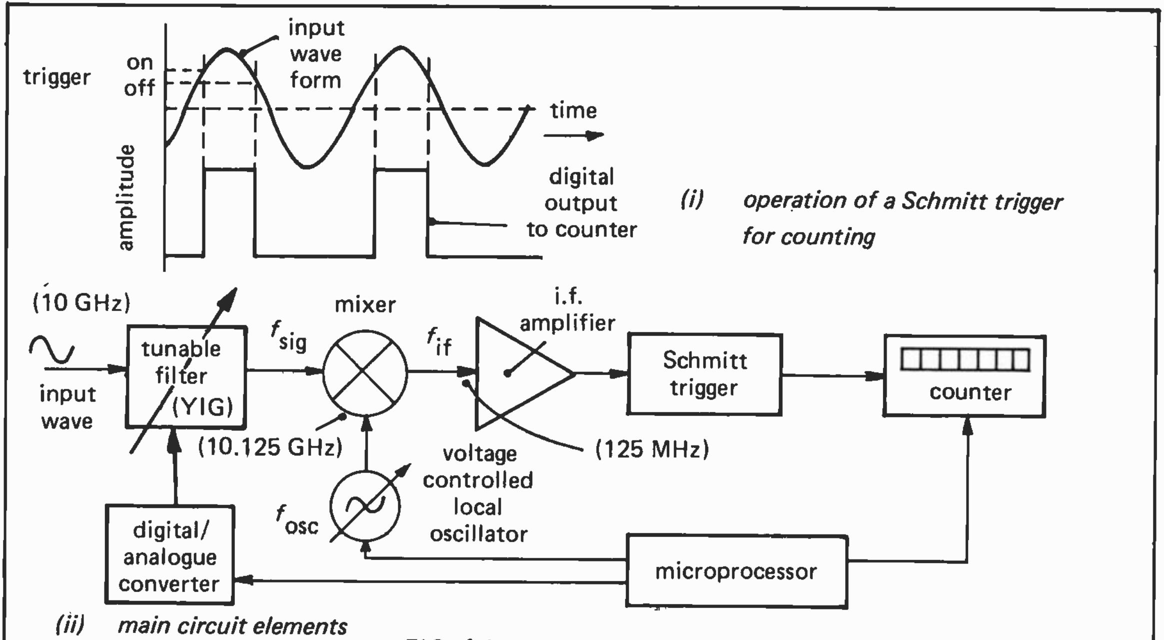
MEASUREMENT

Little can be achieved in research and development without measurement for we then neither know how things are nor what we have done. Most of the measurement techniques nowadays employed have been specially developed for microwaves because of the limitations of standard (lower frequency) equipment. A particular example of this is given by the cathode-ray oscilloscope (c.r.o.) which is unable to display microwave voltage waveforms above a few GHz. In this section we look at just three types of measurement which are essential for most systems. It must be emphasized that what follows is far from complete, we are only scratching the surface of measurement, the intention is mainly to illustrate the differences between "standard" measurement practices and those used for microwaves.

4.1 Frequency

This is probably the most fundamental parameter needed for description of a wave. As mentioned above the standard c.r.o. cannot cope but fortunately recourse may be made to several other methods, for example *cavity resonators* (Sect.2.1.4) and *transmission line resonators* (Sect.2.1.2). These resonate at certain known frequencies which can be adjusted by plungers and are capable of good accuracy.

More modern is a system based on the digital counter, such an instrument normally has microprocessor control. Typically a digital counter may employ a *Schmitt trigger* to convert an incoming continuous wave into a corresponding train of pulses as shown in Figure 4.1(i). The counter section then counts the rate of arrival of the pulses and displays the result. To bring the input frequency down to a value at which the counter can operate, the superheterodyne principle is employed as shown in Figure 4.1(ii). For the benefit of some readers we perhaps first need to recall the superheterodyne technique which is that when a local oscillator frequency (f_{OSC}) is mixed with a



(i) operation of a Schmitt trigger for counting

(ii) main circuit elements

FIG. 4.1 MICROWAVE FREQUENCY MEASUREMENT signal frequency (f_{SIG}) in a non-linear device, sum and difference frequencies ($f_{OSC} + f_{SIG}$) and ($f_{OSC} - f_{SIG}$) are generated and one of these (usually the latter) is filtered off as the intermediate frequency (f_{IF}) so that:

$$f_{IF} = (f_{OSC} - f_{SIG})$$

for example as shown in the diagram in brackets, when a local oscillator frequency of 10.125 GHz is mixed with a signal frequency of 10 GHz, the outcome is an intermediate frequency of 0.125 GHz (125 MHz).

The microprocessor adjusts the tunable band-pass filter (Sect.2.5.3) to select the wave to be measured while rejecting harmonics and other unwanted frequencies. The input waveform and that of a local oscillator controlled by the microprocessor are mixed so that the intermediate frequency is generated as confirmed by the counter. The local oscillator is voltage controlled (e.g. varactor tuned – Sect.3.4.4) so its frequency is known to the microprocessor from the magnitude of the voltage which has to be applied. The microprocessor therefore is able to calculate and display the input wave frequency on the counter. Clearly accurate calibration of the local oscillator is essential and this is subsequently maintained by a crystal controlled reference oscillator (not shown). This is simply the basic system, many other refinements are built in when truly precise frequency measurements are required.

It is worth noting that in frequency we have one parameter of a wave which, unless it meets deliberate frequency changing on its way, is not changed even minutely as it travels through a system. This is more than can be said for most other parameters such as voltage, current, power, attenuation, etc.

4.2 Power

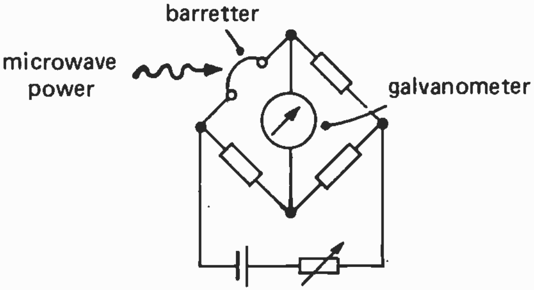
At frequencies lower than microwave we are used to measuring electrical power in terms of voltage, current and resistance. We can appreciate that these quantities are not easily measurable for microwaves when we consider not only the frequencies encountered but also the fact that transmission is often by waveguide where "terminals" tend to lose their meaning. At microwave frequencies therefore power is frequently measured via the change in resistance of a thermally sensitive element.

4.2.1 Barretters

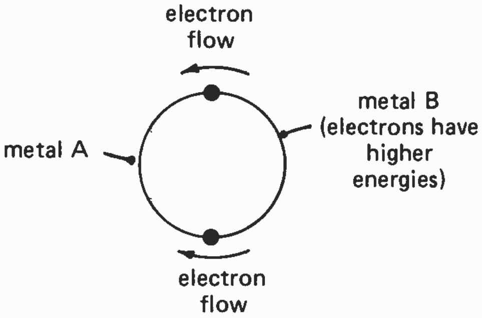
A *bolometer* is defined as an electrical instrument for measuring heat radiation, hence in microwave engineering the term is generally used to describe a resistive, temperature-sensitive element which undergoes a resistance change when it absorbs electromagnetic radiation. One type of microwave bolometer is the *barretter* which is especially useful for the measurement of low and medium powers. This in essence may be a platinum wire of only around one-thousandth of one millimeter in diameter but capable of absorbing microwaves. A thin metal film on a substrate of glass or mica may also be used. The wave or film has a positive temperature coefficient (resistance increases with temperature). Because the wire for example is so fine its resistance change is measurable even with powers as low as 0.01 μW. Changes of some $4 - 10\ \Omega$ per milliwatt of applied r.f. power are typical. The resistance is normally measured by a Wheatstone bridge circuit as shown in Figure 4.2(i). For a wire barretter a cartridge mounting is usually employed not unlike that of a small fuse.

4.2.2 Thermocouples

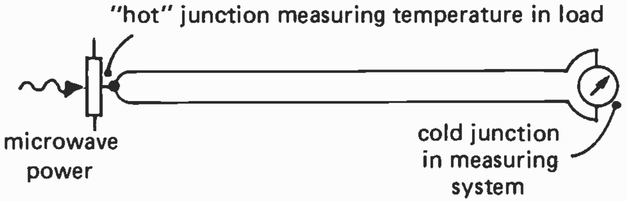
A thermocouple comprises two dissimilar metals (usually wires) joined at their ends and generating an e.m.f. proportional to the temperature difference between the two junctions [the thermoelectric (Seebeck) effect]. If a loop of two dissimilar metals *A* and *B* is set up as in Figure 4.2(ii) then if for example, the extra energy required by the electrons for escape from the surface of metal *A* is greater than that for *B* (the *work functions*), then when in contact more electrons pass from *B* to *A* than in the opposite direction. With both junctions at the same temperature the e.m.f.'s of the two junctions are in opposition and therefore cancel out. If one junction only is heated, cancellation is incomplete and the net e.m.f. increases with the difference in temperature between the two junctions. In a practical thermocouple therefore the "cold" junction is part of the measuring system while the



(i) a barretter in a Wheatstone Bridge circuit

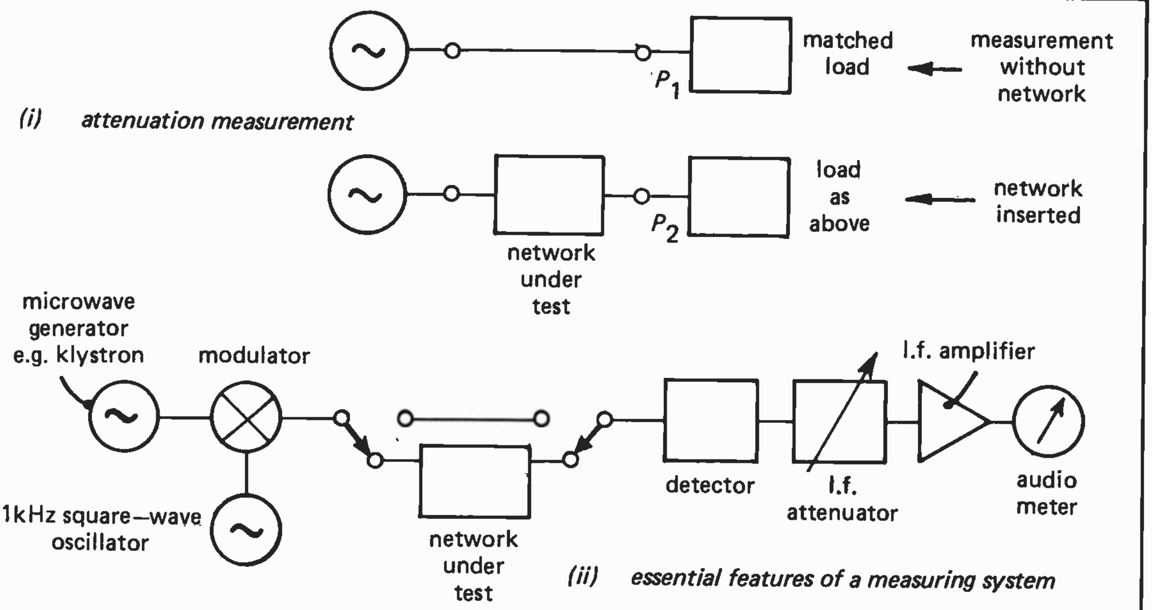


(ii) two thermocouple junctions at the same temperature



(iii) practical measurement of power

FIG. 4.2 MICROWAVE POWER MEASUREMENT



(i) attenuation measurement

(ii) essential features of a measuring system

FIG. 4.3 MICROWAVE ATTENUATION MEASUREMENT "hot" junction senses the temperature of a load in which the microwave power is dissipated [see (iii) of the figure]. The "cold" junction temperature must of course be kept constant. The heating takes time to settle down, hence thermocouple microwave power meters have slow response times, e.g. from a few tenths of a second up to several seconds. However, they have good sensitivity.

4.3 Attenuation

In general attenuation is the reduction in magnitude of signal power as it travels over a transmission path. It is solely due to the resistance present and can be expressed as the ratio between the output power of a network and that at the input. Measurement of the attenuation of a network is usually accomplished by measuring the power (P_1) from a signal source then measuring a second time (P_2) with the network inserted as shown in Figure 4.3(i). The signal source and its load are matched. The network attenuation is then calculated from 10 log (P_1/P_2) decibels.

The bare bones of a practical system are shown in (ii) of the figure. This uses the above substitution method but for convenience of measurement first modulates the microwave signal by, for example, a 1 kHz square wave, thereafter the equipment works at this lower frequency. The signal being used to determine the network attenuation is therefore a series of bursts of microwave power at the test frequency occurring 1000 times a second and it will be seen that with a detector of the barretter type, the measurement at the end of the chain can be at an audio frequency.

This is a circuit in all simplicity, a complete system providing high accuracy also involves isolators, tuners and standing-wave indicators with a microwave frequency measurement system also built in.

Chapter 5

THE MAGIC OF MICROWAVES – COMMUNICATION SYSTEMS

Communication they tell us is the science and practice of transmitting information, all very straightforward when we are together, but when apart, especially a long way, then we expect telephone calls to be cheap and plentiful, television likewise and we even need to learn of the problems of the world "live by satellite". Microwaves play an important part in all of this. But how do we get from *A* to *B* by microwave when *A* and *B* are a long way apart? Certainly not by waveguide as a glance at Figure 2.5 will show, cost and complexity are too great. Circular waveguides may have been used for long distance telephony not so long ago, but these are now eclipsed by the optical fibre. Coaxial cables are less costly but as Section 2.2 shows they give up at frequencies above about 3 GHz.

There is only one transmission medium left and fortunately it is the least expensive of all – the space around us. Hence most microwave communication is by radio. However, although true space does not attenuate a radio wave, the earth's atmosphere and the earth itself do, so we find that how a microwave travels depends greatly on these two factors and according to the wave frequency.

5.1 Propagation

Radio transmission of microwaves is accomplished basically in two different ways. We can see this for ourselves from the television antennas, those for UHF stations (Fig.1.2) point to a local television transmitter, the others for SHF transmissions point up into the sky to a far distant satellite. To appreciate the difference we might usefully start at the lower frequencies and work upwards.

5.1.1 The Ground Wave

The energy radiated by most broadcast radio transmitters (this does not include satellites) consists mainly of a *ground wave* which travels over the earth's surface and a *sky wave* which travels via the atmosphere. The earth is an electrical conductor and as the ground wave moves over it, charges are induced in the ground and these charges travel with the wave. A moving charge is an electrical current, hence power is dissipated and this power can only come from the wave which therefore is attenuated. How far away from the transmitter a useful radio signal can be picked up depends mainly on the transmitting antenna power output, its directional properties and in a rather complicated way on the frequency, the higher this is, the greater the rate of attenuation of the wave.

Above about 1 – 2 MHz the ground wave is attenuated so rapidly that it is of little use except over short distances. This is why at the lower microwave frequencies (say UHF) television distributed by ground wave has a small service area and many transmitting stations are required to cover even a small country (e.g. over 600 for the UK). There are, of course, many other services which employ short distance ground wave communication, especially mobile ones. So we might conclude (with tongue in cheek) that the UHF band (300 MHz – 3 GHz; Sect.1.5) is usable for ground wave propagation but because of ground losses, the SHF band (3 – 30 GHz) is not.

5.1.2 The Sky Wave

Below the microwave range, say up to about 70 MHz, radio communication is possible by using a wave projected upwards by the transmitting antenna into the *ionosphere* and reflected back to earth. In this way the wave effectively skips over the ground instead of travelling in contact with it. The ionosphere which is a layer of ionized atmosphere some 50 to a few thousand kilometres high causes refraction of electromagnetic waves and effectively bends them back to earth. The effect varies with wave frequency and many other factors. It does not really concern microwaves for their frequencies are sufficiently high that the waves pass straight through the ionosphere with little refraction, this is just one of the reasons why satellite transmission is in the SHF range (see also Sect. 5.3.2).

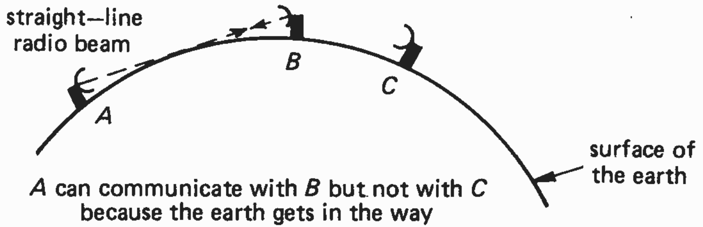
There is also *tropospheric* propagation (up to 10 km high) and this does extend into the microwave band, up to some

30 GHz. It is such a volatile transmission path however that it is not a serious contender for normal microwave communication systems.

5.2 Point-to-Point Microwave Communication

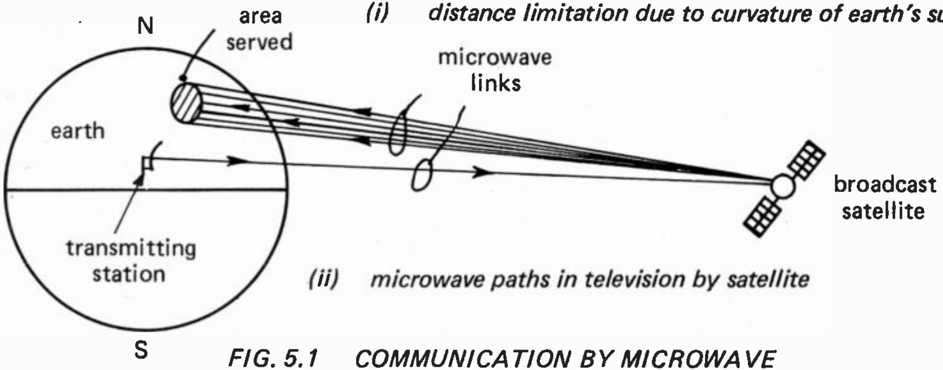
What is generally referred to as a *telephone network* carries much more than ordinary telephone conversations, the network also conveys telegraph, computer, facsimile and other digital data, radio and television baseband signals (*baseband* refers to the range of frequencies initially generated, e.g. the output of a microphone or television camera). It is the avowed intention of the information technology people that these auxiliary services will increase abundantly so perhaps the overall title "telephone network" is not so appropriate nowadays. However, to replace it is not easy. The term *point-to-point* encompasses all the services mentioned, so it is used for the title of this section. This is not to say that other systems such as radar and electronic warfare do not have their point-to-point features but for convenience we agree that here the term refers to whatever happens on a telephone network.

Point-to-point microwave transmission is seen in many large cities in the shape of a very tall tower fitted with microwave antennas. Such for example is the Telecom Tower in London, this houses the microwave transmitters with waveguide feeds to the antennas which are on the outside, each pointing to its partner antenna up or down country. As with light, nothing must block the beam, so the tower is of sufficient height to clear all buildings and other obstacles. The maximum distance tolerable between a microwave transmitter and its receiver is limited by the curvature of the earth as illustrated in Figure 5.1(i) and there is a further limitation caused by mist and rain. At microwave frequencies the wavelength may be of the same order as the diameter of droplets, hence there is some absorption and scattering so allowance must be made for the worst atmospheric conditions. Although point-to-point microwave systems have carried telephony and television channels successfully for many years, high maintenance costs of the antennas and the advent of better technology



A can communicate with *B* but not with *C* because the earth gets in the way

(i) *distance limitation due to curvature of earth's surface*



(ii) *microwave paths in television by satellite*

FIG. 5.1 *COMMUNICATION BY MICROWAVE*

(e.g. optical fibre transmission) are likely to lead to their replacement as time progresses.

Satellite transmission follows the general principles outlined above. Figure 5.1(ii) shows the basic system. It is truly point-to-point on the way up but for television a slightly divergent beam is used when the satellite transmits the signal back to earth in order to cover a certain area, e.g. a small country. There is nothing in pure space to attenuate the wave, but there is some loss through the earth's atmosphere.

5.2.1 Assembling the Package

If calling whatever is to be transmitted by microwaves a *package*, be it analogue or digital and for telephony, television, data, etc., then for transmission over a microwave link it is perhaps obvious that the package can occupy a wide frequency band, even up to some 100 MHz. A band like this can transport over 20,000 telephone conversations at the same time! A single telephone channel needs no more than 4 kHz bandwidth and a colour television signal some 6–7 MHz (*bandwidth* is considered in Appendix 3) so it is obvious that many circuits can be assembled together to form a single package for transmission over a microwave link. This might operate for example through two pairs of dish antennas on the ground as in Figure 2.7(i) or over a satellite link as in Figure 5.1(ii) in which case the down link is also point-to-point.

We touched on the principle of "mixing" two frequencies to provide a third, their sum or difference in Section 3.5. On this technique is based the method of assembling individual channels to form a complete package to be carried on a microwave link. How single telephone channels find their way into a microwave package is explained in essence below. A 4 kHz speech channel mixed with a 104 kHz oscillation will produce amongst others a range of frequencies running from 100 to 104 kHz. A similar speech channel mixed with 108 kHz results in a band 104 to 108 kHz. These two higher frequency channels can be combined and provided that there is no non-linearity, they will not intrude one upon the other just as broadcast radio transmissions do not normally interfere. A *group* of 12 speech channels raised to a band 60 – 108 kHz may further be frequency converted with other groups to higher frequencies still and so on until a package of the required size is formed. The system is known as *frequency division multiplex.*

Digital signals from computers, facsimile machines or from speech which has been converted into digital form are treated differently. Then time division multiplex is used in which each channel is allocated a different time interval for its transmission. All channels are examined in rotation and at each instant of time the signal from only one channel is transmitted.

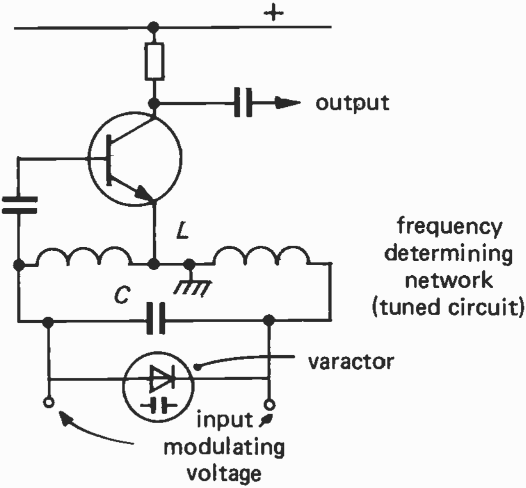
Summing up, a package may contain a mixture of analogue and digital signals, each occupying its own individual frequency band of the width required. The whole package can then be carried by a microwave radio signal.

5.2.2 Modulating a Microwave Carrier

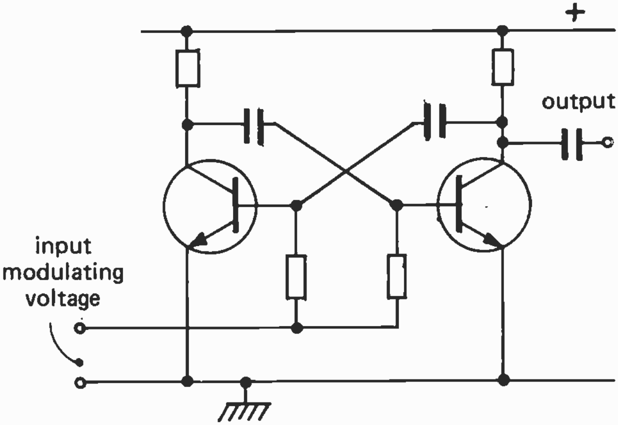
Communication by radio invariably embraces *modulation* (see also Appendix 3) which is the way we impress information onto a carrier wave. Here the carrier wave is usually a sine wave at microwave frequency whereas the modulating wave is at a considerably lower frequency and may be extremely complex, consisting of speech, pulse, television waveforms or a mixture of all.

Of the several different ways of impressing a signal onto a carrier wave, *frequency modulation* (f.m.) is almost invariably used for microwave systems. It has several advantages over its forerunner, *amplitude modulation* (a.m.) especially in that it is less affected by radio path noise. However, for the same input signal a wider carrier bandwidth is required. Happily sufficient bandwidth for large multiplex systems is available in the microwave range.

With frequency modulation the carrier wave (in this case the microwave) is of constant amplitude and its frequency is varied about the mean value in accordance with the amplitude of the modulating wave while the *rate* of variation is at the frequency of the modulating wave itself. Of the many types of modulator in use, the simplest to understand is that of an oscillator circuit tuned by an inductance-capacitance (*LC*) network with the frequency variation arising from changes in the tuned circuit capacitance. Such capacitance variations may be developed by the connection of a varactor (variable



(i) a Hartley oscillator with added varactor



(ii) voltage control of a multivibrator

FIG. 5.2 FREQUENCY MODULATORS FOR MICROWAVES capacitance diode) in parallel with the tuned circuit as shown for a Hartley oscillator in Figure 5.2(i). The modulating voltage is applied to the varactor, hence the net tuned circuit capacitance and therefore the oscillator frequency vary in sympathy.

Slightly more complicated is the multivibrator circuit shown in outline in (ii) of the figure. The levels at which the transistors switch are controlled by the input modulating voltage, hence the output frequency changes accordingly. The circuit may not be designed at microwave frequency but at a lower one, stepped up as required by a frequency multiplier. When integrated circuits are involved, an *operational amplifier* may be used to replace the multivibrator.

5.2.3 Transmission

The modulated carrier, raised suitably in level by a power amplifier is fed via a waveguide to a parabolic antenna for launching into the atmosphere. Over a ground link there is the restriction on path length as illustrated by Figure 5.1(i) and also as mentioned in Section 5.2 there is a signal loss caused by mist and rain. Generally, therefore, "one hop" path lengths do not exceed about 60 km. For system lengths greater than this (and many are), *repeater stations* are used, usually evident as a tall mast with receiving dishes on one side and transmitting dishes opposite. These receive the weak modulated microwave signal, amplify it and then by use of a mixer (Sect.3.5), change the carrier frequency to one slightly different from the received one. The new carrier is amplified and then transmitted onwards. Typically a 4 GHz incoming signal may be re-transmitted on 4.2 GHz. The change in frequency is required because some of the high-level output signal may find its way back to mix with the low-level input signal. If both are at the same frequency, instability is likely to occur. The change in frequency greatly reduces the risk of instability. There may be many repeater stations en route and at each successive one the frequency shift is reversed.

When the microwave signal finally arrives at its destination it is collected over the area of a parabolic receiving dish and reflected into a waveguide. Filtering and mixing produce an intermediate frequency (say, 70 MHz) for further amplification and limiting. Limiting cuts off all wave excursions above a predetermined level so that variations in amplitude are mostly removed, hence reducing noise. This can be done with frequency modulation systems because the *amplitude* of the signal carries no information. The signal is then demodulated so that the original band of modulating frequencies is regained. With a frequency modulated wave a *frequency discriminator* is required. A circuit commonly used firstly expresses the frequency deviation of the frequency modulated signal as a phase-shift, then as an amplitude variation, i.e. the circuit operates as an f.m. to a.m. converter. A conventional a.m. detector then produces an output at the modulation frequency (see also Sect.3.4.1). After amplification the demodulation product (which is a copy of the original package) is divided down by the use of band-pass filters into its component circuits for transmission to the various destinations.

This is for one direction only, e.g. A to B. Many circuits such as those carrying television baseband data are transmitted in one direction only, others, for example, speech circuits are two-way, hence a return circuit (B to A) is also required. With such a system as described above, B to A is identical with A to B and usually exists side-by-side with it. Thus the last repeater station in the A to B direction is also the first in the B to A direction.

5.2.4 Mobile Communications

With point-to-point communication, when one station is mobile a whole host of additional problems arises. No longer is the stability of a pair of wires (or channel in a fibre-optic system) from the telephone exchange to the user's telephone available but instead a radio path. The radio signal has to enter buildings, go up and down hills, find its way through streets, avoid trees and generally cope with the fact that the receiving antenna is low down, could be anywhere in the area and even on the move on a vehicle. Given all this and every one of these conditions has to be met, are microwaves of any use and if so, what range is suitable? We can make some rough estimates.

5.2.4.1 Frequency Considerations

A frequency modulation system is preferred and this in fact requires a greater bandwidth per channel compared with an amplitude modulation system, something approaching 30 kHz is needed, i.e. the channels need to be spaced at least 30 kHz apart (Appendix 3). Each radio channel must be duplex so that conversation in both directions at the same time is possible. Needing perhaps over 600 duplex channels on a system so that most users can be accommodated when they all call or are called together leads to a frequency bandwidth requirement of, say, 40 MHz. The indications are that only microwaves can provide bandwidths of this order, especially if several radio carriers are needed to cover a large area.

Returning now to basic theory, it can be shown that when transmitter and receiver antennas are dipoles, the power delivered to the receiver varies inversely as f^2 where f is the frequency, for example if f rises from 100 MHz to 1 GHz the received power is reduced by 20 dB (more on decibels in Appendix 1). However, it can also be shown that antenna gain varies directly as f^2, hence these two frequency conditions tend to balance out. From this it appears that the higher microwave frequencies with their abundance of bandwidth can be employed – but there are overriding factors, two which perhaps place the greatest limitations are:

(i) *buildings and trees* – the signal has to find its way inside a building. Experimental work on the attenuation suffered by an electromagnetic wave when passing through a brick wall shows that it may be a few decibels at frequencies at low megahertz but can be as much as 40 dB at 3 GHz, clearly at this latter frequency getting through a wall is difficult, at even higher frequencies almost impossible. We have more evidence of this in the fact that at the higher still frequencies of light it is manifestly impossible. Tree losses also increase with frequency and above 1 GHz dense trees are a major obstruction to a wave.

(ii) *rainfall* – we can guess from the effects of rainfall on light that microwaves, although of somewhat lower frequencies will also be affected. It has been estimated that in heavy rain an attenuation of 0.01 dB/km at 2 GHz rises to around 10 dB/km at 20 GHz. Even moderate rainfall attenuates a 20 GHz wave at a rate of 3 dB/km.

Putting (i) and (ii) together indicates that the higher microwave frequencies are unsuitable for mobile communications, a compromise between the conflicting requirements appears to limit the use of the microwave spectrum to the UHF band, i.e. to frequencies below about 3 GHz.

5.2.4.2 The Cellular System

We have seen from Section 5.1.1 that microwaves at the lower gigahertz frequencies are capable of serving no more than a relatively small area unless very high transmitter powers are used. In open areas things are not so bad but in heavily built-up urban areas the received signal can be as much as 60 dB lower. Accordingly the *cellular system* has been developed with considerable success, but it must be admitted that this depends heavily on microprocessors and large-scale integration, in fact until these came along the service could never have got off the ground. An idea of the method can be gained from Figure 5.3. The service area is divided up into a number of *cells* (hence the description of the system), cells of the same number all working on the same group of microwave frequencies. The drawing shows the cells to be of similar area but of course in practice they may differ widely in size. The number of different cells is shown as 7 but the actual number employed depends on the type of area to be served. The connections are shown between non-mobile and mobile subscribers, this of course does not rule out connections between two mobile subscribers. It must be emphasized that this is the basic idea only, many variants exist and many more will be developed.

Each cell contains fixed transmitting and receiving antennas, the transmitter radiating sufficient power (several watts) for the complete single cell and supplying all cells of the same number as shown. Interference between cells working at the same frequency is minimized because of the rapid attenuation of the wave with distance. Throughout each cell a

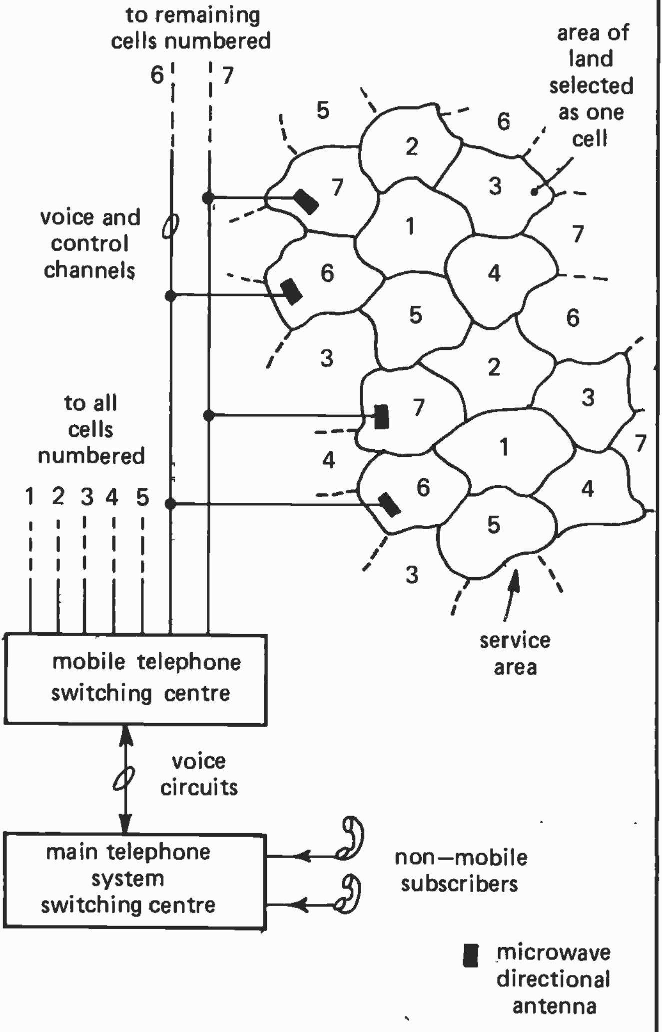


FIG. 5.3 TYPICAL CELLULAR SYSTEM LAYOUT

satisfactory service should be available. However, when a vehicle moves across the boundary from one cell to an adjoining one its own transmitter and receiver are automatically realigned to the frequencies of the new cell. This of course is no mean feat. Firstly the controlling computer must know the location of each mobile telephone. It is the signal-to-noise ratio (Appendix 2) on the radio path which is the main determinant of the intelligibility of the signal it carries, this must therefore also be checked, with the system able to switch the mobile receiver frequency to the new one as quickly as possible, say within less than 40 ms. All this requires a separate data link with the mobile system, the same link is also used for setting up and terminating calls.

5.2.4.3 Reducing Signal Fluctuations

If we consider a vehicle moving along a street dominated by buildings, it is clear that a microwave signal being received or transmitted will suffer uncommonly large amplitude variations including short-duration fades. Signal variations of greater than 20 dB may be expected. Fortunately the majority of fades to almost a complete loss of signal are of very short duration, the actual duration decreasing with the speed of the vehicle for example, fades of 10 – 20 ms at 4 mph falling to only a few milliseconds at 20 mph. Even though these times seem short, overall such variations must result in an unsatisfactory signal-to-noise ratio, hence without correction the degradation in voice transmission could be serious, for data transmission, disastrous.

A considerable improvement is obtained by using a *diversity* system. This employs two or more cell antennas for both transmit and receive with directional characteristics as required. It is possible then for the receiving equipment in either direction to switch to the antenna providing the highest signal-to-noise ratio or alternatively the outputs of the antennas can be combined on the assumption that, even though spaced only a wavelength or more apart, they are unlikely to experience signal fades at the same instant.

Voice processing might also be employed. Telephone microphones vary in their sensitivities, talkers vary far more in their voice levels and how far away from the mouth the handset is held. The system therefore has to accept low electrical speech levels from weak voices at large speaking distances to high levels from loud voices close to the microphone – this could easily be a range of up to 60 decibels. Such a range would create difficulties in the modulation of the frequency modulation carriers, low voice signals resulting in a reduction in channel clarity, high ones experiencing *clipping.*

Much can be done to improve the signal-to-noise ratio and hence ensure good overall transmission quality, especially by:

(i) *pre-emphasis and de-emphasis* – by this method the high voice frequencies are raised in level before modulation in the transmitter (pre-emphasis), then reduced in the receiver by the same amount (de-emphasis) so that overall a flat response is obtained. By increasing the higher frequencies during pre-emphasis the signal-to-noise ratio over the channel is improved and this is maintained during de-emphasis because the reduction in noise is the same as that of the signal.

(ii) *companding* – is another system for improvement of the signal-to-noise ratio over a channel. The voice frequencies generated by a talker firstly meet a *volume compressor* which provides more gain for low amplitude signals than for high amplitude ones. Low-level signals are therefore less affected by noise on the channel because of their increased amplitudes. At the receiving end a *volume expander* restores the voice signal to its original range of amplitudes.

Both (i) and (ii) may be used together with the result that much of the variability of the microwave link is ironed out and users experience satisfactory transmission quality. These few notes which are mainly restricted to how it is possible to talk over a mobile system are only sufficient to give us an inkling of the complexities involved. The control system also is indeed a frightening collection of integrated circuits for it must at least (i) know where every vehicle is, (ii) determine whether it is available to receive a call, (ii) transmit that call, (iv) respond to a mobile originated call, (v) switch to a different transmitter when a vehicle moves across a cell boundary,

(vi) handle the transfer of calls to and from the main telephone system.

5.3 Broadcast Communication

Broadcast as opposed to point-to-point simply implies a wide dissemination of information. As far as microwaves are concerned television reigns supreme in the broadcast field. Here we consider only the two types of television broadcast systems generally entitled "terrestrial" (the wave travels over the ground) and "satellite" (the wave travels down from a high-flying satellite).

5.3.1 Television (Terrestrial)

Section 5.1.1 shows how a ground wave is attenuated and suggests that at frequencies higher than UHF the ground wave has little to offer. Accordingly some terrestrial television is broadcast in the VHF band (30 – 300 MHz) but most is at a higher frequency at the lower end of the UHF band (300 MHz – 3 GHz). However, there is some help from experimental work which has shown that wave energy lost to the ground decreases with height above the ground to the extent that at heights of several wavelengths the wave suffers very little. In fact the strength of the ground wave transmitted from a half-wavelength dipole varies directly with $h_t \times h_r$ where h_t is the height of the transmitting antenna and h_r that of the receiving antenna. Thus some improvement in the received signal is obtained from siting television antennas high on roofs. Nevertheless because of other obstacles such as hills and buildings and the fact that as frequency rises smaller objects such as trees and leaves create absorptions, we can sum up with the conclusion that in terms of the microwave range terrestrial television is limited to the lower end [see Fig.1.2(ii)].

At the transmitter the power amplifier is most likely to be a klystron (Sect.3.2.1) with its output fed to the antenna by a waveguide (Sect.2.3). The main type of antenna used for receiving is well known and easily recognized as a Yagi (Sect. 2.4.1) with a coaxial lead to the television receiver. Indoor or set-top receiving antennas may also be used in strong signal areas.

5.3.2 Television (Satellite)

Without microwaves there would be no satellite television at all so perhaps our first consideration should be the fact that microwaves can reach up to satellites whereas waves of lower frequency cannot. This involves looking a little more closely into the effect the ionosphere has on waves of different frequencies. The ionosphere (first mentioned in Sect.5.1.2) arises from the ionizing influences of solar radiation. This acts on particles in the atmosphere and supplies energy to electrons enabling them to escape from their parent atoms which are then left positively charged and known as positive ions. A free electron may subsequently combine with a neutral atom or molecule to form a negative ion. The concentration of ions in the ionosphere is extremely variable, changing with day and night, season, height and sun-spot activity. Satellite television is in the SHF band (3 – 30 GHz) and what is so helpful about the ionosphere is that it does not refract these frequencies and return them to earth as it does certain lower frequencies.

Surely none of us has escaped the standard school physics experiment of dipping a stick into water and finding that it appears bent. This is because the velocity of a wave in the denser of the two media (air and water) is lower and it is easily shown that this results in the wave being deflected or, as we say, *refracted.* We assess this by the *refractive index* (μ) which is simply the ratio of the wave velocities in the two adjacent media. When $\mu = 1$ there is no refraction.

Ionization changes within the ionosphere can be considered as adjacent media of different densities, hence as a wave progresses through the ionosphere it can experience continual refraction and under certain circumstances be deflected back to earth. This is the opposite of what is required for satellite transmission for it is essential that the wave gets through to the space beyond where there is nothing to interfere with its progress no matter how long the journey. The mathematical formula which sorts this out is more than a little complex but by making several assumptions (e.g. that there is no magnetic field present and there are no particle collisions) it is possible to indicate what happens as frequency rises. Clearly μ must be as near to 1 as possible for the waves to go straight through:

$$\text{Refractive Index, } \mu = \sqrt{1 - (kN_{\max}/f^2)}$$

where f is the frequency under consideration and $N_{\max}$ is the maximum electron density. Clearly from this, as f increases, μ approaches unity and any refraction of the wave is reduced. Of course there are so many variables around (especially $N_{\max}$) that the formula can only be taken as a guide yet it is sufficient to indicate how microwaves are usable for satellite television whereas lower frequencies are not. Generally in Europe the frequencies employed are around 12 GHz.

To see how microwaves function in satellite television let us follow them on their journey from camera to television set. The baseband frequencies from the camera and microphone occupy a bandwidth from 0 to 6 – 8 MHz. This band frequency modulates the microwave carrier (say around 12 GHz) resulting in a microwave signal of up to some 27 MHz bandwidth centred on the carrier frequency. The increase in bandwidth follows from the need to accommodate sufficient of the harmonics produced in the modulation process. After modest power amplification (a few kilowatts) by a klystron or travelling wave tube, the signal is conducted by waveguide to the transmitting antenna. The antenna will be basically of parabolic shape [Fig.2.7(ii)] aimed at the satellite. On its journey through the earth's atmosphere the microwave signal suffers some absorption, refraction and scattering but the transmitter power is such that a signal of ample strength arrives up there.

Most satellites are in geostationary orbits (stationary with respect to earth), positioned at nearly 36,000 km vertically above the equator and travelling at just over 11,000 km/hour. At this height and speed the satellite completes one orbit in exactly the same time as the earth rotates once, hence to an observer on earth the satellite appears to be stationary. The signal is received on a parabolic dish, amplified and frequency changed so that up and down microwave frequencies differ slightly to avoid instability. The signal is then power amplified by a travelling wave tube and fed to the transmitting antenna by waveguide. This antenna is usually a single parabolic or several working together, but for certain purposes a *horn* antenna may be used, a type which can be likened to a waveguide opened up into a horn shape, it of course does not have the same high directivity of the parabolic.

Satellites are powered by the sun so the transmitter output power is limited, generally to no more than 200 watts. The wave polarization will be vertical, horizontal or circular (Sect. 1.4). On the journey down to earth there may be diffusion and absorption by rain, mist and clouds in the lower layers of the atmosphere which could account for a loss of several decibels. The strength of the signal picked up by a ground antenna (usually parabolic) varies with its size and dishes of from some 30 cm up to 1.6 m are common, the larger sizes being required for reception from more distant or less powerful satellites. Microwave signals received by the dish are directed into a *feed horn* coupled to a *low noise block converter* (LNB). This device is a microwave frequency changer which accepts the incoming band (or block) of signals and changes them to a similar band but centralized on a lower frequency so that coaxial cable can be employed between the antenna and the satellite receiver instead of a waveguide (the cost of such an ungainly contraption would be prohibitive). The feed horn and LNB must be adjusted to match the incoming wave polarization and the output is likely to be between 1 and 2 GHz, which is still in the microwave range. However, a second frequency changing stage in the satellite receiver results in a carrier of, say, 70 MHz, so at this point the microwave story ends. The 70 MHz signal is demodulated to regain the baseband frequencies which then control the television receiver.

Chapter 6

THE MAGIC OF MICROWAVES – RADAR

We might consider radar (an acronym from RAdio Detection And Ranging) to have first seen the light of day in February 1935 when Sir Robert Watson-Watt produced a notable memorandum entitled "Detection and Location of Aircraft by Radio Methods". The basic principle on which this was founded is the same today, that is if an aircraft or other object is irradiated with electromagnetic energy, a small amount is re-radiated and can be detected. Furthermore, if the energy were to be transmitted in short high-power pulses then by measuring the time elapsing between transmitting the outgoing pulse and receiving the reflected one, the distance away of the reflecting object could be calculated. Very many refinements have been added to this basic idea as the years have passed, perhaps the most fundamental one being the use of highly directional antennas so that not only can the distance away of the object be measured but also its bearing.

Although the earliest experiments were conducted at about 12 MHz, microwave technology has taken over since its very short wavelengths allow the use of small antennas and reflections are more sharply defined. It is also found that signals may be returned from other than metal objects. Generally radars operate at frequencies from 600 MHz to 40 GHz (50 cm to 7.5 mm).

6.1 Reflection of Radio Waves

Section 2.1.2 considers the reflection of electromagnetic waves, albeit in a much simplified fashion. We normally think of transmission lines as lengths of metallic conductors along which a wave progresses. Now we must expand our thinking to consider the atmosphere also as a transmission line for it is in fact a medium through which an electromagnetic wave is transmitted. It has its own characteristic impedance (Z_0) and other constants. For the more technically inclined the characteristic impedance of free space is given by $\sqrt{(\mu_0/\epsilon_0)}$ where μ_0 is the magnetic constant of free space ($4\pi \times 10^{-7}$ H/m) and ϵ_0 is the electric constant (8.854×10^{-12} F/m). This gives the value of Z_0 as 376.7 ohms. The atmosphere has its own characteristic impedances not greatly different from this but varying with its make-up.

When radio wave energy therefore meets a discontinuity in the atmosphere (i.e. anything which has not the same characteristic impedance as the atmosphere), some energy is reflected back along the transmission line, in other words it is returned to the transmitter. The likelihood of any solid mass in the atmosphere having the same characteristic impedance as the atmosphere itself is practically zero, hence radar relies on the principle that an object will reflect some of the wave which hits it.

6.2 The Radar Range Equation

Although much of what is considered in this section may to some be obvious, this interesting equation is developed to provide useful reminders of the difficulties with which the radar designer has to grapple. That so much success has been achieved is demonstrated by the fact that nowadays a missile can use its radar to catch up with another one. The theory is at its best only a guide because so many of the factors affecting the distance over which a particular radar system can operate are exceedingly variable. The concept of the *isotropic antenna* may seem to some a little strange so we take this first step by step.

The isotrophic antenna is a purely theoretical idea developed for assessing the usefulness of practical antennas. The isotropic is considered to be a point source radiating equally in all directions, hence if, for example, a transmitting antenna radiates an electromagnetic wave power of P_t watts then at any distance R metres away this power can be considered as being spread over the surface of an imaginary sphere of radius R. Hence the *power flux density* (p.f.d.) R metres away is $P_t/4\pi R^2$ watts per square metre (W/m^2) as developed in Figure 6.1. The *gain* of any practical transmitting antenna in a given direction is then given (usually in decibels) by:

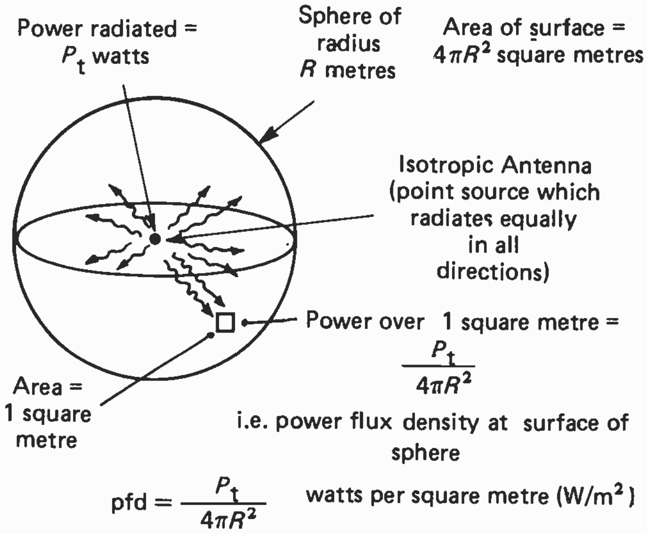


FIG. 6.1 RADIATION FROM AN ISOTROPIC ANTENNA

$$G_t = \frac{\text{power radiated by the antenna}}{\text{power radiated by an isotropic antenna supplied with the same transmitting power}}$$

Equally the gain of a receiving antenna, G_r is measured as:

$$G_r = \frac{\text{signal power delivered by the antenna}}{\text{signal power delivered by an isotropic antenna for the same transmitted power}}$$

Consider a *target* at a range R in free space. As shown above the p.f.d. at R will be

$$\frac{P_t}{4\pi R^2}$$

for an isotropic antenna but

$$\frac{P_t G_t}{4\pi R^2}$$

for a practical antenna of gain G_t.

Now suppose that the target intercepts this power over an effective area of σ square metres and re-radiates it back over the same distance R to the receiving antenna:

$$\text{total power re-radiated (reflected)} = \frac{P_t G_t}{4\pi R^2} \times \sigma .$$

The p.f.d. at the receiving antenna is derived by assuming that the target is now a transmitter, hence:

$$\text{p.f.d. at receiving antenna} = \frac{P_t G_t \sigma}{(4\pi R^2)^2} \text{ W/m}^2 .$$

Suppose that the effective area of the receiving antenna is A_r and that its gain relative to the isotropic is G_r, then:

signal power output, P_r of receiving antenna

$$= \frac{P_t G_t \sigma}{(4\pi R^2)^2} \times A_r \text{ watts} .$$

A_r is related to the gain by the basic antenna formula:

$$A_r = \frac{G_r \lambda^2}{4\pi}$$

i.e.

$$P_r = \frac{P_t G_t \sigma}{(4\pi R^2)^2} \times \frac{G_r \lambda^2}{4\pi} \text{ watts}$$

and if, as is most likely to be the case, the same antenna is used for both transmission and reception, $G_t = G_r$ (= G), hence:

$$P_r = \frac{P_t G^2 \sigma \lambda^2}{(4\pi)^3 R^4} \text{ watts .}$$

There will be a minimum value for P_r for the receiving signal to be usable, say, $P_{r(\min)}$, hence by rearranging:

$$R = \left[\frac{P_t G^2 \sigma \lambda^2}{(4\pi)^3 P_{r(\min)}} \right]^{1/4} \text{ metres.}$$

Note the variables:

(i) the formula assumes free space, atmospheric path loss is not included;
(ii) $P_{r(\min)}$ is controlled by system noise which is notorious for its variability;
(iii) σ is perhaps the most variable quantity of all as we might expect. For a large jet aircraft head on σ might be around 10 but side on as much as 100 times greater. Small aircraft show an even greater variation.

Many other variables are also present so we must conclude that accurate calculation from this simple formula is not possible, but it is useful as a guide. Nevertheless two important features the formula does bring to notice are that:

(i) R varies as the fourth power of the transmitter, hence large increases in P_t are required to increase R even moderately, e.g. doubling P_t increases R by less than 20%.
(ii) Technically R increases as the square root of the wavelength. It would appear therefore that the range is increased by lowering the frequency, perhaps below microwave. However, changes in λ have a considerable effect on the size and design of the antenna so a compromise has to be adopted.

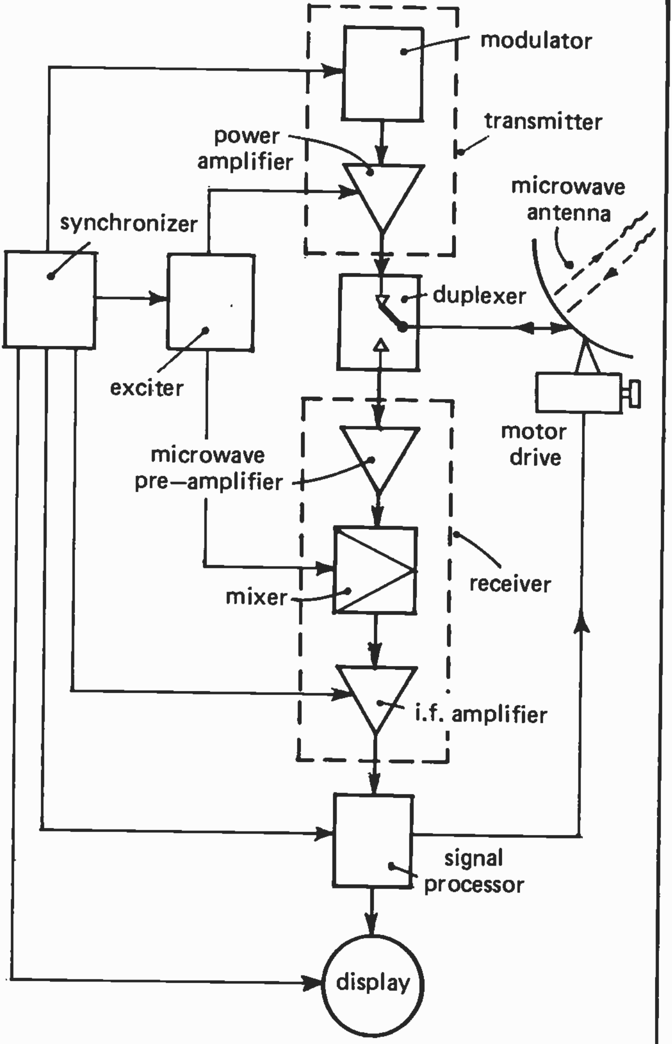


FIG. 6.2 A PULSED RADAR SYSTEM

6.3 Elements of a Radar System

Radar might be defined in a few words as a system for ascertaining simply the presence or more likely the direction and range of objects such as aircraft, ships and buildings. A typical basic pulsed radar system capable of satisfying these requirements is illustrated in Figure 6.2. Many modern techniques such as are used in navigation, weather prediction, astronomy, early warning air and sea surveillance and even missile destruction may all be highly sophisticated systems yet they are merely extensions based on this moderately simple arrangement. Needless to say, computer technology abounds in plenty. The system shown in the figure is, of course, static, when it is built into a moving object such as an aircraft or a missile then things really do get complicated.

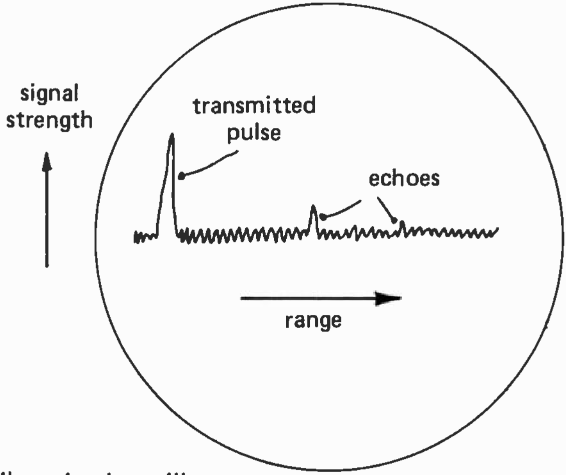
It all starts with the *synchronizer*, the function of which is to control the timing of the whole system. To send out a radar pulse the synchronizer causes the transmitter *modulator* to send out a high voltage pulse to the power amplifier which may be a klystron (Sect.3.2.1) or for the higher powers, a magnetron (Sect.3.3.1) or sometimes a travelling-wave tube (Sect.3.2.2). At the same time the synchronizer triggers the *exciter* into applying a microwave drive signal to the power amplifier, the net result being a high power microwave pulse fed from the power amplifier to the *duplexer* via a waveguide.

The duplexer is basically a send/receive switch, essential when a single antenna is employed for both transmission and reception. It also has the function of protecting the receiver from the high power of the transmitter. It must operate within the time between transmission of the outgoing pulse and reception of the echo. The one shown in Figure 6.2 is depicted as a mechanical switch but it is obvious that the switching speeds required are beyond the capabilities of a mechanical device. Several different types of duplexer are in use, a ferrite circulator as discussed in Section 2.5.2 is perhaps an obvious choice, such devices can handle the very high transmitted power levels and provide some 20 – 30 dB attenuation between transmit and receive. A type also commonly used is the *gas-tube* which is inserted within the waveguide system. On receiving the low powered signal passes through the tube unattenuated but the high power of the transmitter causes the gas to ionize and its resultant low resistance short-circuits the r.f. energy and prevents it from affecting the receiver. Two gas-tubes in a hybrid arrangement are required for a complete balanced duplexer. More recently high power microwave diodes are coming into favour as replacements for gas-tubes. Via the duplexer therefore a single pulse of microwave energy is radiated. In the drawing the antenna is shown as being steered by a motor drive, equally a stationary antenna may be used with the beam steered electrically.

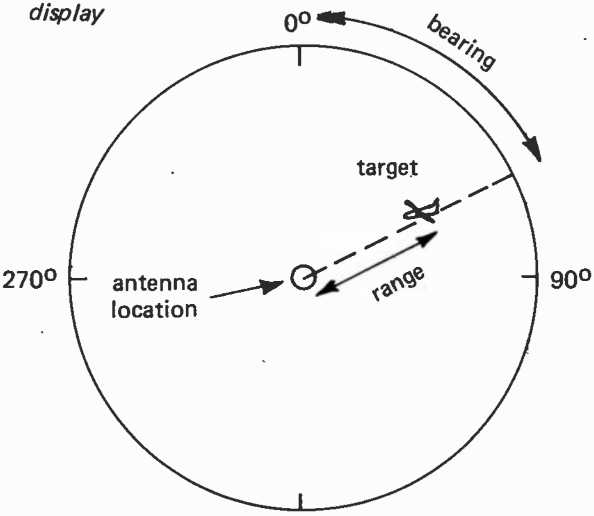
The reflected signal is received by the antenna, it is directed to the receiver via the duplexer. It is first amplified by a low-noise microwave preamplifier and then, using the superheterodyne principle converted to an intermediate frequency (Sect. 3.5) for transmission to the *signal processor.* The local oscillation for the mixer is supplied at the correct time by the exciter. This is the ideal situation but usually many other weak signals and noise (*clutter*) also arrive from the antenna so the main purpose of the signal processor is to discriminate against the unwanted signals, obviously a complex computer operation. The output is then fed to the display system, in its simplest form a cathode-ray tube which gives a range-only presentation as sketched in Figure 6.3(i) but is more likely to be of PPI (plan-position-indication) form indicating both range and bearing as shown in (ii), here the time base continually rotates to correspond with the position of the antenna at any instant. Clearly the repetition frequency and length of the pulses must be such that the echo does not return after the following transmitter pulse for on a display as in Figure 6.3(i) it might get lost in the transmitted pulse display or if arriving later still may be mistaken for a short range echo of the second cycle.

6.4 Navigation and Electronic Warfare

We put these two sub-titles together because in general they are both based on an extension of the radar principle so that it applies to moving objects, i.e. either the observer or the target or both are in motion. Two examples only are: (i) the control of aircraft from the ground; and (ii) one missile seeking out another. To do no more than examine the basic



(i) simple oscilloscope display



(ii) PPI display

FIG. 6.3 RADAR DISPLAYS

principle is as far as we dare go, otherwise things become too complicated and lengthy. However, what follows should be enough for an appreciation of how the *Doppler effect* is used with microwave technology to measure speeds of distant objects.

The Doppler effect is noticed as the apparent change in pitch of a sound wave from a moving body as it passes by (first considered by Christian Doppler, an Austrian physicist). The effect is not, however, limited to sound waves but occurs similarly with electromagnetic waves including those of light where slight changes in colour arise.

Consider the system in its simplest form as that of a radar transmitter of frequency f_t radiating continuous waves (i.e. the pulse technique of Sect.6.3 is not used). The echo is detected and mixed with a low-level sample of f_t. It is perhaps obvious that if both the transmitting station and the target are stationary the echo will be at the same frequency as f_t. However, if one or both of these is moving the echo will have a different speed imposed upon it and there will be a difference in frequency between the transmitted and received signals (the *Doppler shift*). From the basic physics equation the radar one follows:

$$\text{Doppler shift, } f_d = f_t \left[\frac{(c+v)}{(c-v)} - 1 \right]$$

where c is the speed of propagation of the transmitted wave (f_t) and v is the radial velocity of the target. The formula can be adjusted to account for movement of both transmitter and the target.

Through a little juggling and approximation a very simple formula arises:

$$f_d \simeq \frac{2v}{\lambda}$$

where λ is the wavelength.

Let us take for example an aircraft approaching a radar station at a speed of 200 m/s (approximately 450 m.p.h.) and assume that in air $c = 3 \times 10^8$ m/s. For a transmitted frequency at 1 GHz, f_d is calculated from both formulae to the nearest whole number as 1333 Hz, so clearly the shortened formula is sufficient as a guide. Since in the practical case f_d is easily measurable, v can simply be derived from $(f_d \lambda)/2$.

Many radar techniques are in use based on the general radar principles outlined in the preceding sections together with the Doppler shift technique so that radar technology has now reached the point where practically any civil or military requirement can be met.

Chapter 7

THE MAGIC OF MICROWAVES – HEATING

That microwaves are used for heating is known to all, especially the cooks among us. In trying to understand the process, however, we ought first to understand something about *heat* but when we do inevitably the term *energy* confronts us. Both terms abound in everyday conversation yet we understand so little about what these two quantities really are. Certainly we have no real knowledge of what energy is and the dictionary gives little help for it explains energy away as "the ability of matter or radiation to do work", i.e. in terms of what it can do rather than what it is. Trying to explain heat is also difficult for generally we find that heat is described as a *form* of energy.

Held in the hand a cricket or golf ball is no threat, but if we get in the way of one while it is in flight, it can do nasty things to us. What has changed? Nothing except that it has been given a copious supply of Nature's mysterious quantity, the energy of motion – kinetic energy.

Atoms and molecules in all matter are continually in a state of vibration. They therefore possess kinetic energy. If heat is added the motion is increased so the kinetic energy increases also. On the other hand if heat is removed the motion is reduced. Thus one way of describing heat is that in matter it can be perceived as the kinetic energy of atomic and molecular motion. Temperature is a measure of the average of this energy. Put simply, temperature does not measure the amount of heat in an object, it measures its hotness.

An electromagnetic wave has energy which can be transferred to the vibrating particles in matter (atoms, electrons, molecules) through interaction with its electric and magnetic fields. The energy is kinetic hence the vibrations increase and from our explanation above, heat is "generated". Infra-red radiation has long been used in this respect but now the next-of-kin, the microwave does the job more efficiently. We examine the production of heat in a *dielectric* in more detail in the following sections.

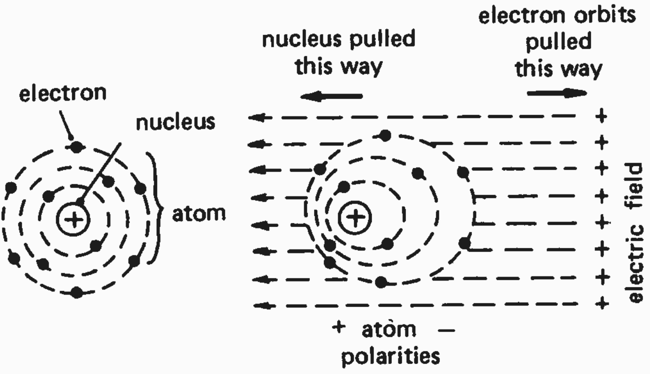
7.1 Charge Displacement

We whose initiation included Ohm's Law and the subsequent formulae for power (E^2/R, I^2R) may be excused for thinking about electrical heating in terms of a resistance wire glowing hot. But microwave heating has added another dimension, it actually heats a dielectric which from our knowledge of capacitors should never occur to any great extent. After all, the definition of a *perfect* dielectric is a material which accepts no energy from an applied electric field and therefore it must have infinite resistance and cannot heat up. Practical capacitor dielectrics do have some charge carriers although in most cases, very few. An applied voltage therefore results in such a small current that electrical resistance heating is negligible.

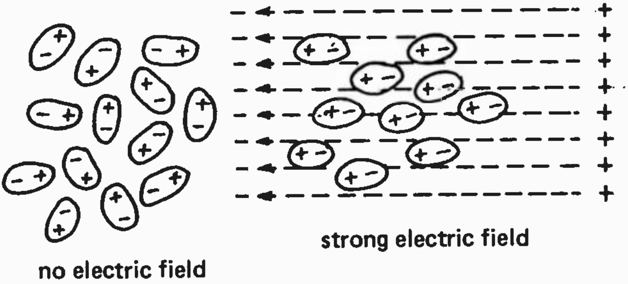
Now we meet a different form of heating altogether, no longer relying on the Ohm's Law type of free charge carrier movement but in fact operating on all the atoms or molecules or both (for convenience let us call these *particles*) of the whole material. Very briefly, when a suitable dielectric is placed in an electrical field, charges within the particles are *displaced* and when the applied field is alternating the displacement follows the field polarity changes. To do this energy must be absorbed from the field and it is then dissipated as heat. When charges within a particle are displaced, the particle is said to be *polarized.* Different kinds of polarization can arise but fortunately heat seems to be generated whatever happens. We look at the least complicated form only, it will be sufficient to gain a good understanding of microwave heating.

7.2 Polarization

When a dielectric is subjected to an electric field, the particles (atoms or molecules) become elongated in the direction of the field, the electrons being pulled one way, the nuclei in the opposite direction as sketched in Figure 7.1(i). The charges within each particle are now unbalanced to the extent that the ends of it are oppositely charged – the particle is said to be *poled*, it is called a *dipole*, the process being known as *polarization.* Furthermore, each particle may be aligned in the direction of the field, hence giving rise to a redistribution of



(i) creating an atomic dipole



(ii) dipoles responding to an applied electric field

FIG. 7.1 DIELECTRIC POLARIZATION

charge within the material as shown diagrammatically in (ii). Losses, appearing as heat occur in the dielectric because energy is absorbed from the electric field in aligning the particles. If the electric field is reversed, the dipoles reform in the opposite direction and if the field alternates at some high frequency, the dipoles continually readjust, the energy dissipation resulting in a continuous generation of heat – from within the material. The magnetic field of the electromagnetic wave has an insignificant effect because there is nothing in most dielectrics it can work on.

Even for this simplified explanation, the mathematics can get out of hand so we take the key dielectric heating equation only. From this the important features can be seen:

$$\text{power generated within a dielectric} = 2\pi f E^2 \epsilon_0 \delta \text{ watts per unit volume}$$

where E is the voltage applied at a frequency f, ϵ_0 is the electric constant (permittivity of free space) and δ is the effective *loss factor*. This latter term is not an easy one to get to grips with unless one has studied dielectrics in depth, it is technically the product of the power factor and the relative permittivity of a dielectric. For a given electric field the heat generated is proportional to the power factor. However, its name gives us a clue for clearly the greater the overall loss within a dielectric, the more the power which will be absorbed from an electromagnetic wave. It is a summation of the various different losses and is both frequency and temperature dependent.

7.3 Microwave Heating Frequencies

Most importantly the formula shows that the heating effect (power generated) varies with the frequency and with the square of the applied voltage. Considering now the practical condition, it is evident from the formula that too high a voltage is likely to lead to breakdown, better therefore to rely on a high value for f and this puts its value in the microwave region. Overall control is needed nationally since high powers at microwave frequencies are in danger of propagating electromagnetic radiation which may interfere with television, medical or other scientific equipment. A list of microwave heating frequencies has therefore been agreed.

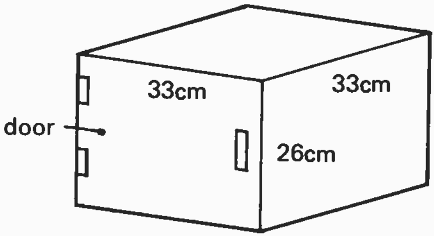
It is clear from the formula that the product $f \times \delta$ is the factor determining the wave voltage required for a given degree of heating. As an example, most foods contain water which it so happens has the highest value for δ (about 16 at 15°C at 3 GHz – for most dry materials δ is less than 0.5), this most often leads to the choice of 2450 MHz (2.45 GHz) for domestic ovens, 896 MHz is also used although less frequently.

7.4 A Typical Microwave Heating System

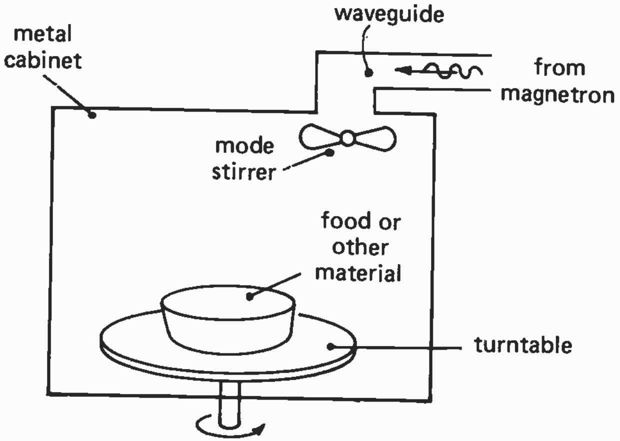
Because most readers have experience of domestic microwave cookers it is propitious that this is the type we examine, safe in the knowledge that basically industrial types are little different. The microwave energy is radiated from an antenna system into the free space within a closed cabinet whereupon it is absorbed by materials (dielectrics) having a high loss factor, in the case of the domestic oven such a material is mainly the water contained within food. A typical domestic cooker closed cabinet might have the dimensions shown in Figure 7.2(i) with the general construction as sketched in (ii).

The closed cabinet is constructed of a conducting material such as aluminium or stainless steel. These metals have almost zero loss factors so absorb no power, in fact practically all of the microwave reaching them is re-radiated. The cabinet may be considered as a rectangular waveguide closed at both ends. Microwave energy fed in is continually reflected by the metal walls in all directions and standing waves are set up with their nodes and antinodes (Sect.2.1.2).

In the majority of microwave cookers the microwave power is provided by a fan-cooled magnetron (Sect.3.3.1) and it is fed into the cavity via a short length of waveguide. The standing wave patterns would normally cause uneven heating, this is minimized by the *mode stirring* fan which consists of slowly rotating blades which project the microwave energy in different directions to vary the field patterns. The item to be heated is placed on a motor-driven turntable, again to distribute the heat evenly. The door and its locks are so designed that power is disconnected from the magnetron when the door is opened. Door sealing is such that energy radiation into the open air is prevented. The whole cooker may be rated at some



(i) typical cabinet dimensions



(ii) typical construction

FIG. 7.2 A MICROWAVE COOKER

1250 watts for an electromagnetic wave power of 650 watts.

In industry the same principles may apply but the ovens are naturally of larger dimensions, e.g. for defrosting food in bulk.

Tunnel ovens are employed in which the material moves along the tunnel on a conveyor belt (e.g. for curing rubber extrusions, drying leather). Small but powerful ovens are used for heating resins and other chemicals.

7.5 Microwave Diathermy

Never having considered ourselves to be dielectrics it may come as a surprise to find that microwave diathermy (from Greek, *through heat*) operates on the principles outlined above for microwave cookers. In this medical treatment microwave energy is transmitted into the skin from a distance of a few centimetres. The aim is to heat deeply placed muscle, not of course to cook it, hence precautions are built in so that over-exposure to the microwaves is impossible. Both patient and physiotherapist wear wire-mesh goggles to exclude radiation from the eyes.

Appendix 1

DECIBELS

It may have had its origin in the era when scientists were beginning to measure the intensity and other characteristics of sound waves. It was then discovered that homo sapiens seemed to have logarithmic ears for the intensity of sound had to be increased many times before it was considered to be only twice as loud. A logarithmic system of measurement seemed to be the answer and it was soon found that using logarithmic units to express power, current or voltage *ratios* has two distinct advantages: (i) large numbers are reduced and so become more manageable; and (ii) in a complex system with a large number of circuits, each contributing a gain or loss, calculation of the overall power, etc., ratio by multiplication is unwieldy. By expressing each ratio in a logarithmic unit, addition takes the place of multiplication, a more manageable process altogether. Hence the *bel* came on the scene (after Alexander Graham Bell, the Scottish-American inventor) which is simply the common logarithm (to the base 10) of the power ratio. This gives rather low figures so an offshoot soon appeared, the *decibel* (symbol dB) which is one-tenth of one bel. If the power, current or voltage input to a circuit is given by P_1, I_1 or V_1 respectively and the power output by P_2, I_2 or V_2, then the attenuation (or gain):

$$n = 10 \log \frac{P_2}{P_1} = 10 \log \frac{I_2^2 \times Z}{I_1^2 \times Z} = 20 \log \frac{I_2}{I_1} \text{ dB}$$

but note that I_1 and I_2 must be measured in the same terminating impedance, Z. Similarly:

$$n = 20 \log \frac{V_2}{V_1} \text{ dB}$$

and again the voltages must be measured in the same impedance. (Note however that we often deliberately go astray when quoting the voltage gain of an amplifier by using the above formula while ignoring differing input and output amplifier impedances. No problem if we realize what we are doing and tell others.)

If gains and attenuations are being considered together, then signs become necessary. If P_2 is greater than P_1, there is amplification and n is positive, conversely if P_1 is the greater, there is a loss and n is negative. Suppose $P_1 = 1$ mW and $P_2 = 0.5$ mW, obviously there is a power loss so n should work out to be negative:

$$\text{attenuation in dB } (n) = 10 \log (P_2/P_1) = 10 \log 0.5$$

$$= 10(-0.3010) = -3.01 .$$

Absolute values can be quoted in decibel notation provided that a *reference* or *zero* level is stated or known. As an example, the reference level of a quantity such as signal power or sound pressure is chosen and this is given the decibel value of 0. The reference level is usually indicated by an added letter to the symbol. One commonly used in electronics is one milliwatt for which the symbol now becomes dBm. Thus a power level of 100 mW may be expressed as +20 dBm because a power gain of 20 dB on 1 mW results in 100 mW. No reference level need be quoted because it is indicated by the "m". Equally –30 dBm is the same as 1 μW. Another commonly used reference level is one volt for which the symbol used is dBV and there are several used for sound and noise levels [e.g. dB(A)] .

Note that the practice of quoting noise levels directly in decibels used by the general public is technically wrong, but it works.

Three conversions worth remembering are:

a power ratio of 1.25 : 1 is equivalent to 1 dB (approx.)
a power ratio of 2 : 1 is equivalent to 3 dB (approx.)
a power ratio of 10 : 1 is equivalent to 10 dB.

These can often be used to avoid switching on the calculator, for example:

$$27\ \text{dB} = 10\ \text{dB} + 10\ \text{dB} + 3\ \text{dB} + 3\ \text{dB} + 1\ \text{dB},$$

representing a power ratio of 10 × 10 × 2 × 2 × 1.25 = 500, (the exact answer is 501.2 showing that the quick method suffers very little from loss of accuracy).

Appendix 2

SIGNAL-TO-NOISE RATIO

This is a most important measure of the capacity of a telecommunications channel to carry information. Noise is inevitable, it is the uninvited and generally unwanted guest. It is defined as any spurious electrical disturbance occurring within the effective frequency band of a circuit. Noise is of many forms, here is a reminder:

(i) *thermal noise* is always present due to the random movements of free electrons within a conductor. The effect increases with temperature;
(ii) *semiconductor noise* arises from the random fluctuations in all of the currents especially by the random arrival and departure of charge carriers moving across a junction;
(iii) *crosstalk* occurs in line systems, it is the unwanted voltage induced in one channel due to signals in another channel;
(iv) *radio interference noise* arises when signals are received from a carrier having a frequency close to that of the wanted signal;
(v) *electrical apparatus interference* injects noise into a system through rapid changes in current or from sparking, so generating harmonics;
(vi) *natural noise* from the earth's atmosphere or from the galaxy affects radio systems.

The signal-to-noise ratio (s/n) is the ratio of the wanted signal power to the noise power accompany it, i.e.:

$$\mathrm{s/n} = \frac{\text{wanted signal power}}{\text{noise power}}$$

or if to be quoted in decibels (Appendix 1):

$$\mathrm{s/n} = 10\log_{10}\left[\frac{\text{wanted signal power}}{\text{noise power}}\right]\text{dB}$$

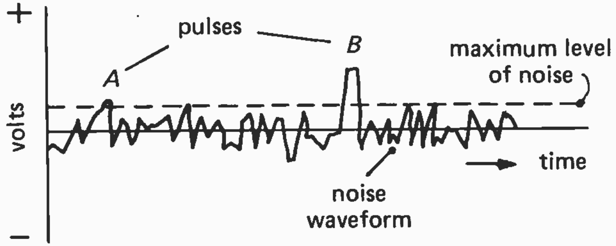


FIG. A2. SIGNALS IN NOISE

For any particular system the signal-to-noise ratio required depends on the type of signal and its use. A high quality music circuit may require a signal-to-noise ratio of some 60 dB whereas an ordinary telephone call may be considered satisfactory with a ratio of only 15 – 20 dB.

A graphical demonstration of the importance of the signal-to-noise ratio is given in Figure A2.1 which shows how in a transmission channel carrying digital information, the signal-to-noise ratio must be reasonably greater than 1 to avoid errors, i.e. a pulse not being recognized or alternatively a noise spike being mistaken for a pulse. It will be seen that the pulse at *A* has little chance of being detected above the general noise level whereas the one at *B* is better off because equipment can be arranged to accept only voltages significantly above the maximum noise level. It is evident therefore that what matters most is the *degree* by which the signal exceeds the noise and this is given by the signal-to-noise ratio.

Further proof of how noise can degrade transmission quality has been demonstrated by C. E. Shannon (an American mathematician). His well published formula is:

$$C = W \log_2 (1 + s/n)$$

where C is the information capacity of a channel of bandwidth W, showing clearly that the amount of information a channel can successfully carry (C) improves as the signal-to-noise ratio increases.

Appendix 3

MODULATION

Taking it for granted that we can generate a high frequency wave and project it into the surrounding atmosphere with the help of an antenna, the question arises as to how it can carry *information* with it. By information is meant such things as the electrical waveforms of speech, music, data, facsimile, television and the like. Clearly on its own the high frequency wave has nothing to offer for the only information which can be gained from it is whether it is on or off. We make it carry information by changing its form in some way at the frequency and amplitude of the originating information, known as the *baseband*. The wave is then appropriate known as a *carrier wave* and the process of imprinting the baseband on it is called *modulation*.

Almost invariably the carrier starts off as a sine wave (Sect.1.2) and it is a golden rule that when a sine wave is altered in any way, then frequencies other than that of the wave are generated. Modulation therefore results in a *band* of frequencies usually centred on the carrier. The width of this band is important for it determines how many radio transmissions can be packed into a given range of radio frequencies. We examine the two most widely used ways in which a carrier wave is modulated so that the particular choice for microwave working can be appreciated.

Amplitude modulation (a.m.) has the amplitude varied in exact proportion to that of the baseband so that effectively the tips of the carrier wave trace out the shape of the baseband waveform. The band of frequencies produced by amplitude modulation has a width twice that of the highest modulation frequency, e.g. if the highest modulating frequency in the baseband is 5 kHz and it modulates a 1000 kHz carrier, the result is a band of frequencies extending from 995 to 1005 kHz. Theoretically a second transmission carrying a similar baseband would have to be on a carrier of 990 or 1010 kHz to avoid overlap with its consequent interference – in practice a little more spacing is allowed to make sure.

Frequency modulation (f.m.) is the important one here for as Section 5.2.2 shows, f.m. is the method preferred for most microwave systems. Technically, compared with a.m. it is a more complex system. Instead of changing the amplitude of the carrier on modulation, its frequency is varied. The *rate* of frequency variation is according to that of the modulating frequency while the *degree* of variation is proportional to the amplitude of the modulation.

In practice to conserve bandwidth a limit is set on the maximum frequency deviation ($\Delta f_{(\mathrm{max})}$) and if we consider a single modulating frequency f_{m} then the mathematics applicable to such an operation show that:

maximum bandwidth requirement for f.m. wave

$$= 2(\Delta f_{(\mathrm{max})} + f_{\mathrm{m(max)}})$$

which for high quality f.m. radio broadcasting with $f_{\mathrm{m(max)}}$ of 15 kHz and $\Delta f_{(\mathrm{max})}$ set at 75 kHz results in a bandwidth requirement of 180 kHz, much greater than that for an a.m. system of (2 × 15) = 30 kHz.

Bandwidths of 180 kHz cannot be accommodated on the medium wave broadcast band (5 stations would fill the whole band) hence f.m. broadcasting is usually on carrier frequencies upwards of 30 MHz.

In all broadcasting the signal-to-noise ratio (Appendix 2) is especially important for if it is poor the silent background essential in sound and music broadcasting is lost, television develops white spots and graininess and digital transmissions produce errors. This is where f.m. scores over a.m. Most interference signals produce amplitude modulated waves, an a.m. system naturally reproduces these, an f.m. system does not. The price to be paid, however, is in the greater bandwidth requirement. Bandwidth is less of a problem at microwave frequencies, hence although more complicated, f.m. is generally preferred mainly because of its superior signal-to-noise ratios and its requirement of lower transmitter powers.

Index

Notes

Notes

Notes

BERNARD BABANI BP312

An Introduction to Microwaves

- ☐ Take away microwaves and much of modern society collapses - what would we do without our colour television for example? An incentive therefore for those of us so minded to understand the subject perhaps a little better than we already do. This is an introductory book and no background in microwave technology is needed, although some elementary electronics experience is assumed. The mathematics required are little more than ordinary algebra.

- ☐ Organization of the book is straightforward. Brief reminders and technical explanations of those facets of electronics important for an appreciation of microwaves are followed by chapters on microwave generators and amplifiers. This leads on to the practical uses of microwaves in communications generally, mobile communication and television. Finally we look at radar and heating, of course not forgetting to include some notes on the domestic microwave cooker.

- ☐ A book, not for the expert, but for all others who are interested in electronics and feel there is more to microwaves than just cooking.

£3.95

ISBN 0-85934-257-3

00395

9 780859 342575