The precision attainable in the measurement of distance by radar means is very high and it was the use of such methods over accurately surveyed baselines which first drew attention to the fact that the velocity of electromagnetic waves is some 16 kilometres per second greater than the previously accepted value. Two recent determinations, one of which is based on the use of radar techniques, have substantiated this higher value (299,792 kilometres per second) for the velocity of propagation in free space. Activity in many other branches of physical science is similarly being stimulated as the application of radar methods of experiment and instrumentation becomes more widely known.

The successful reception of radar echoes from the Moon by the United States Army Signal Corps in January 1946 is noteworthy as the first step in sounding the universe, a step only made possible by the extraordinary instrumental developments which are an essential part of radar. The application of the same general techniques has opened up an entirely new field of astrophysical investigation, the study of the Sun and the stars by means of the radio-frequency radiation of thermal and electro-dynamic origin which they emit. This is the province of Radio Astronomy, a new branch of science which has already produced something of a revolution in astronomical ideas and which may well lead to a reconsideration of the philosophy upon which our theories of the universe are based.

CHAPTER II

FUNDAMENTALS

1. Components of a Radar Set

THE components of a simple radar set are shown in schematic form in Fig. 4. The sequence of events originates in the

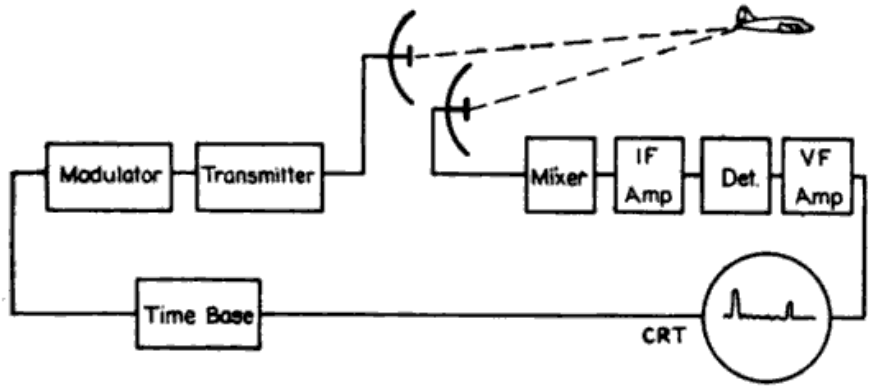


Figure 4.—Components of an elementary radar set.

modulator which produces a series of short DC pulses at the chosen recurrence frequency. These are converted into radio frequency pulses in the transmitter and radiated into space from the transmitting aerial. Some of the energy leaks back to the receiving aerial whence it is conveyed to the receiver, amplified and made to cause a vertical deflection of the cathode ray tube spot. This first signal is known in radar terminology as the " direct pulse." The greater part of the radiated energy travels outwards and that part impinging on hills, ships, waves, aircraft etc. is reflected, giving rise to echo signals in the receiver which arrive with delay times corresponding to the time of travel of the wave to and fro. For clarity in presentation only one echo is shown in Fig. 4.

In order to distinguish the various returned echoes, a horizontal deflection of the cathode ray tube spot, the time base, is superposed giving the characteristic trace depicted in the

figure. This trace is repeated each cycle so that it appears stationary to the eye. The range of the target is then obtained by measurement of the displacement of this echo signal from the main pulse. The bearing is obtained by rotating the aerial in azimuth to give maximum signal strength.

Fig. 5 shows schematically a second equipment which represents a more modern type of set. It has a single aerial

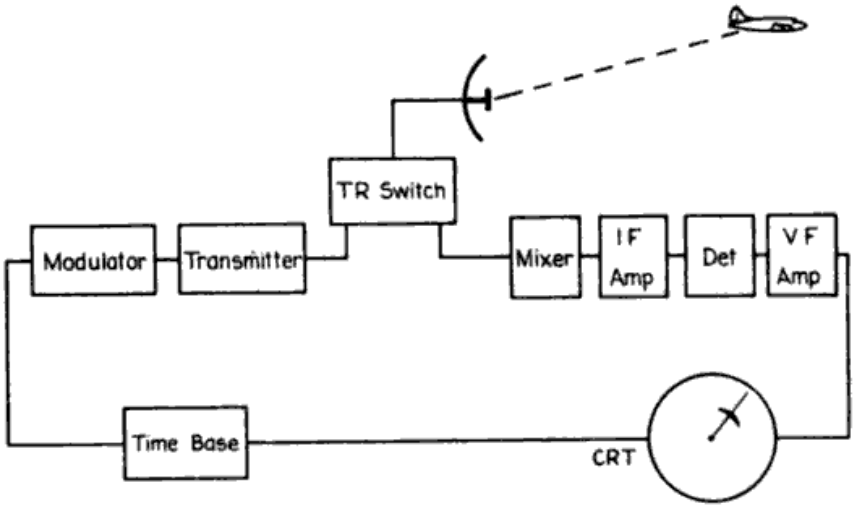


Figure 5.—Components of a typical radar set.

for transmitting and receiving and an elegant form of display known as the Plan Position Indicator (PPI).

In order to transmit and receive on the same aerial, a special duplexing component known as the TR switch is used. This consists of an extremely quick-acting switch which disconnects the receiver during the transmitted pulse and cuts off the transmitter for the remainder of the time.

In the PPI display the time base is radial and starts from the centre of the tube. Instead of causing a deflection, an echo causes the spot to brighten at the appropriate range. The aerial is made to rotate continuously and the time base rotates in synchronism with it. As the beam comes on to a target, the spot brightens, traces a small arc of a circle and disappears again as the beam goes off the target. In this way each echo gives rise to an elongated bright spot at the appropriate range and azimuth. In order to give a more or less

permanent pattern on the face of the cathode ray tube, a long-delay phosphor is used which has a decay period of the same order as that of the aerial rotation. The PPI display is most spectacular when used in an airborne radar equipment looking at the ground below. A picture of the countryside is then drawn on the screen as shown in Plate I.

The arrangements shown are not exclusive; timing may originate in a unit other than the modulator, the aerial may be required to search in elevation as well as in azimuth, and so on. But the two figures do represent arrangements of real radar equipments and serve to illustrate the interrelation of radar components with a minimum confusion.

2. The Echoing Process

The fundamental equation for the intensity Φ of a wave at a distance d from a transmitter of power P and aerial gain G in a given direction is

$$\Phi = \frac{PG}{4\pi d^2}.$$

When a wave is incident on a target a certain amount of energy is intercepted. This is partly absorbed and partly scattered, but for most radar targets the scattering predominates. The directional pattern of the scattered radiation is quite complex and varies radically with slight changes in orientation.

We specify the backward scattering or echoing power of an object in terms of its "echo cross section" σ for a particular orientation and polarisation of incident and scattered waves. If the object is in a plane wave of intensity Φ_o, the intensity of the scattered wave in the backward direction Φ_s at a large distance d is given by

$$\Phi_s = \frac{\Phi_o \sigma}{4\pi d^2}. \qquad (1)$$

Consequently we may associate $\Phi_o\sigma$ for the echo wave with PG for the primary wave and consider $\Phi_o\sigma$ as the product of two factors, total power scattered and directional gain of the scattered power.

In the case of a non-absorbing object large with respect to a wavelength the total scattered power is $\Phi_o S$ where S is the projected area. The case of a large sphere is particularly instructive because it scatters equally in all directions, so that $\sigma = S$. Table 1 gives values of σ for some common objects.

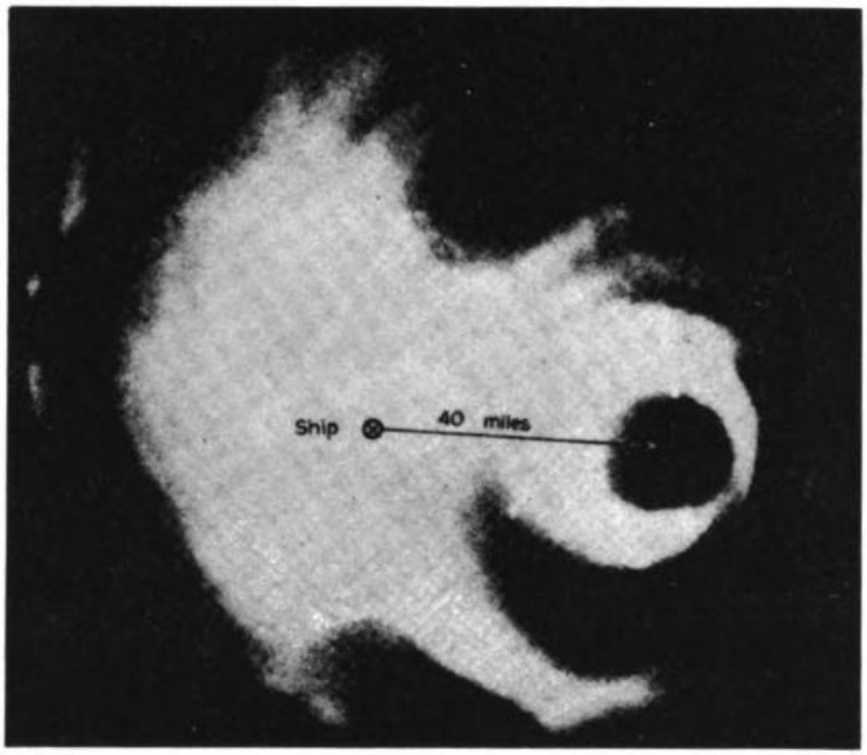
Objects which, for their size, have unusually high values of σ are tuned half wave dipoles and large flat plates normal to the direction of propagation. The former depend on intercepting a large amount of energy relative to their size; the latter on concentrating the energy intercepted. The enormous echo from a large flat surface is restricted to a very small range of angles about the normal. If, however, a ray is reflected in turn from three planes at right angles, the internal corner of a cube, the emergent ray is returned in the same direction as the incident ray and the arrangement gives a large echo over a considerable range of angles.

In military practice the targets of most interest are aircraft, ships, and topographical features. In each case these are large and of complex shape so that the directional diagram is also complex. If the target moves, this lobe pattern changes continually causing the received echo to undergo deep fading which is characteristic of the target.

Reported measurements of the mean values of σ for aircraft show great diversity. However a value of from 1 to 10 square metres, independent of wavelength in the range 3 to 200 centimetres, yields sensible results when used to calculate the performance of radar equipments.

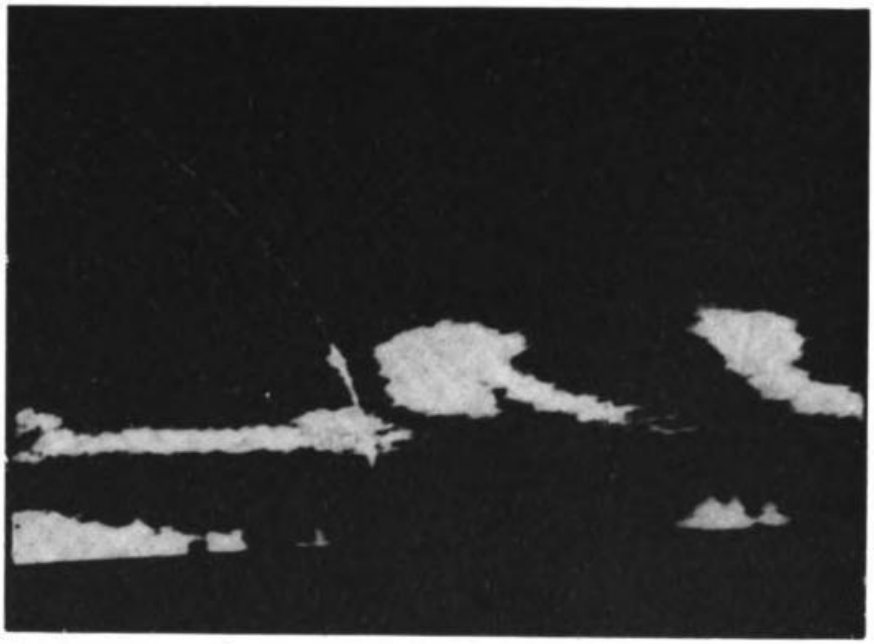
In the case of ships, practical problems are complicated by the fact that the ship extends over a region of grossly varying field. An empirical treatment[1] based on the hypothesis that, on the average, the ship structure is complex enough to be treated as a "white body" in the optical sense, appears to give a fairly good approximation. At close ranges where the echoes are not reduced below those for free space, values of σ

[1] M. V. Wilkes and J. A. Ramsay, "A theory of the performance of radar on ship targets," *Proc. Camb. Phil. Soc.*, Vol. 43, 1947, pp. 220-31.



(*a*) 10-centimetre radar PPI display. Rain associated with a typhoon in the Pacific.

[*By courtesy of Royal Navy.*



(*b*) 3-centimetre radar range-height display of strata of heavy cloud, presumed containing large drops, near Cambridge, Massachusetts, U.S.A. The towering structures to the right of the picture lie between 6 and 8 miles out.

[*By courtesy of Radiation Laboratory Massachusetts Institute of Technology.*

PLATE II.—Examples of rain or cloud echoes.

see p. 37]

TABLE 1—ECHO CROSS SECTION σ OF PERFECTLY CONDUCTING GEOMETRICAL OBJECTS

Object	*Size*	*Orientation*	*Range of Validity*	σ
Dipole (unloaded)	$\lambda/2$	Parallel to electric field	—	$0{\cdot}86\lambda^2$
Sphere	radius a	—	$a \gg \lambda$	πa^2
Curved surface	Radii of curvature ρ_1, ρ_2	Measured where ray normal to surface	ρ_1 and $\rho_2 \gg \lambda$	$\pi\rho_1\rho_2$
Flat plate	Area S	Normal to ray	Dimensions $\gg \lambda$	$\frac{4\pi S^2}{\lambda^2}$
Cylinder	Length l, Radius a	Axis normal to ray	Dimensions $\gg \lambda$	$\frac{2\pi}{\lambda} al^2$

CONTENTS

entire surface, A is equal to the area independently of the wavelength. In practice, due to non-uniform illumination, A is always less than the area, the factor being about 0·6 for common aerials and dropping to very low values for aerials designed to have peculiar diagrams.

For aerials of similar design, equation (3) indicates that the gain for a fixed aperture is inversely proportional to the square of wavelength. In the case of broadside aerials this is associated with a beam width given roughly by λ/a radians where a is the width of the aperture in the relevant direction.

Now let

P = transmitter power delivered to aerial (watts),
p = received power delivered to receiver (watts),
p_o = minimum detectable signal (watts),
G = aerial gain (G_T transmitting, G_R receiving),
A = aerial absorption cross section (A_T transmitting, A_R receiving) (square metres),
Φ_o = intensity of primary wave at target (watts per square metre),
Φ_s = intensity of echo wave at receiving aerial (watts per square metre),
σ = target echo cross section (square metres),
d = target distance (metres),
d_o = ultimate range on particular target (metres).

In free space the primary wave intensity at the target is given by

$$\Phi_o = \frac{PG_T}{4\pi d^2}$$

the scattered wave intensity at the receiving aerial by

$$\Phi_s = \frac{\sigma \Phi_o}{4\pi d^2}$$

and the power delivered to the receiver by

$$p = A_R \Phi_s.$$

Hence

$$p = \frac{A_R \sigma P G_T}{(4\pi d^2)^2}.$$

It is convenient to use this equation to specify ultimate range

d_o of a radar equipment for a particular target. Then d becomes d_o, p becomes p_o, and in addition we use equation (3) to eliminate either A or G, giving

$$d_o^4 = \frac{P}{p_o} \frac{A_R A_T \sigma}{4\pi\lambda^2} \text{ or } \frac{P}{p_o} \frac{A^2\sigma}{4\pi\lambda^2} \tag{4}$$

or

$$d_o^4 = \frac{P}{p_o} \frac{G_T G_R \lambda^2 \sigma}{(4\pi)^3} \text{ or } \frac{P}{p_o} \frac{G^2\lambda^2\sigma}{(4\pi)^3} \tag{5}$$

where in each case the second equation refers to the usual single aerial case.

These are the basic free space radar equations.

It is noteworthy that, because of the twice repeated divergence of energy in radar, d_o varies as the fourth root of P/p_o. Consequently great increases in power pay small dividends in increased range. Aerial aperture is a more important range controlling factor, d_o being proportional to the linear dimensions for similar aerial types.

The variation of ultimate range with wavelength requires careful consideration. When the limit to aerial gain is set by the physical size of the aerial, equation (4) applies and d_o is inversely proportional to $\sqrt{\lambda}$. In certain cases however the aerial is required to have a specified polar diagram so that G, not A, is fixed. Equation (5) then applies and d_o is, in contrast, directly proportional to $\sqrt{\lambda}$.

These equations suggest that a useful figure of merit for the sensitivity of a radar set is the calculated value of d_o for unit echo cross section. It is particularly apt because the effective echo cross section of an aircraft appears to be of the order of one square metre, so that values of d_o calculated in m.k.s. units give approximately correct ranges to aircraft.

Numerical Values

In these equations all quantities concerning the radar equipment are measured and specified except for the minimum detectable signal p_o. This depends on the noise generated in the receiver and on display factors which usually permit detection of a signal about equal to the receiver noise power referred to the input terminal. Table 2 gives these quantities

TABLE 2—CHARACTERISTICS OF TYPICAL RADAR SETS

	Air Warning Types (Land Based)				*Airborne Types*		
Wavelength (centimetres)	150[1]	150[1]	10	3	150	10	3
Peak power (kilowatts)	10	150	700	80	10	100	35
Receiver noise power (watts)[2]	2×10^{-14}	2×10^{-14}	2×10^{-13}	2×10^{-13}	1×10^{-13}	4×10^{-13}	7×10^{-13}
Aerial gain	100	100	10,000	17,000	2	340	1,300
Aerial size (feet)	10 × 20	10 × 20	25 × 8	16 × 3	Dipoles	2·5 (diam.)	1·5 (diam.)
Calculated ultimate range on a target with echo cross section of 1 square metre (miles)	30	60	130	50	3	12	9
Observed range on bomber aircraft (miles)	120	180	180	60	4	10	8
Observed range on fighter aircraft (miles)	70	180	100	40	3	6	3

[1] Observed ranges nearly doubled by earth reflections.
[2] Variations due chiefly to different bandwidths associated with different pulse lengths.

for a few typical radar sets, p_o being estimated in the manner indicated above, together with calculated values of d_o and observed ultimate ranges of aircraft.

Coverage

For a stationary aerial and a given target, the ultimate range in different directions varies owing to the directivity of the aerial. The locus of points at which the target gives a just detectable signal is a surface in space known as the " coverage diagram " of the system.

This diagram is geometrically similar to the polar diagram of the aerial expressed in field strength, not power. This is readily seen since, in a direction in which field is reduced by the factor a in Fig. 6, echo field is reduced in the ratio a^2 and

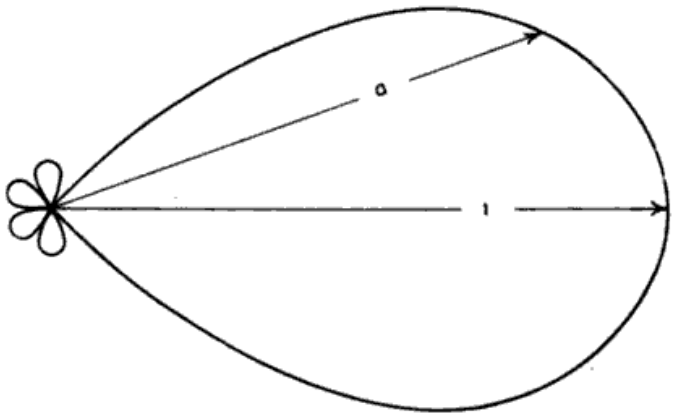


Figure 6.—Coverage diagram of an equipment. This is geometrically similar to the polar diagram of the aerial expressed in volts.

echo power in the ratio a^4. Consequently range is reduced by the factor $(a^4)^{\frac{1}{4}}$ which equals a. The coverage diagram is simply a special contour of a family which gives equal echo strengths and also equal primary fields at the target.

It is clear that the requirement of a specific coverage diagram for a given target defines the aerial gain and hence for a given wavelength the power ratio P/p_o. From equation (5) it follows that this ratio is inversely proportional to λ^2. This result is at first surprising but is related to the greater area and absorption cross section of longer wave aerials for the given gain.

This power ratio is not exactly related to volume or cross

sectional area of a coverage diagram but a very useful approximation exists in the case of a so called " rectangular " aerial. In this case

$$\frac{P}{p_0} \propto S^2 \qquad (6)$$

where S is the area of the section of the coverage diagram cut by a plane through the aerial parallel to one or other side.[2]

This concept, presented for a stationary aerial, can be extended to cover aerials which scan by inserting factors to cover the loss of sensitivity arising from the scanning process. A useful application is to the case of an air warning radar set with an aerial which rotates about a vertical axis. In order to detect aircraft at appreciable angles of elevation the vertical section of the diagram must be wide. If it is too wide the aerial gain and hence ultimate range will be unduly decreased. Equation (6) indicates the actual relationship, S being the vertical section of the coverage diagram. If, for example, the vertical coverage is to be everywhere doubled by adjustment of the vertical aerial diagram alone, the horizontal cover or range will be halved in order to maintain the vertical section constant. Alternatively if we wish to maintain the ultimate range by increase of power, the required increase will be proportional to S^2, or 4 times.

4. Normal Propagation

We shall now consider the departures from free space propagation which occur in practice due to the presence of the earth.

The earth gives rise to a ground reflected wave which reacts with the direct wave and produces an interference pattern in the region above the horizon. Below the horizon there is a shadow region into which some energy passes by diffraction. The field above the horizon may be evaluated fairly accurately from " ray theory " but this method fails in the shadow region. Fortunately radar signals, unlike one-way signals, are usually beyond the limit of detection in this latter region so that most cases can be treated without invoking the full wave theory.

The ray theory concept is that the field at a distant point is the resultant of a direct ray and one or more rays reflected

[2] The diagram has the form $f_1(\theta).f_2(\phi)$, where θ and ϕ are angles with respect to the perpendicular sides of the aerial.

from the earth according to the laws of geometrical optics. In the case of transmission over a smooth surface such as the sea there is a single reflected ray as indicated in Fig. 7.

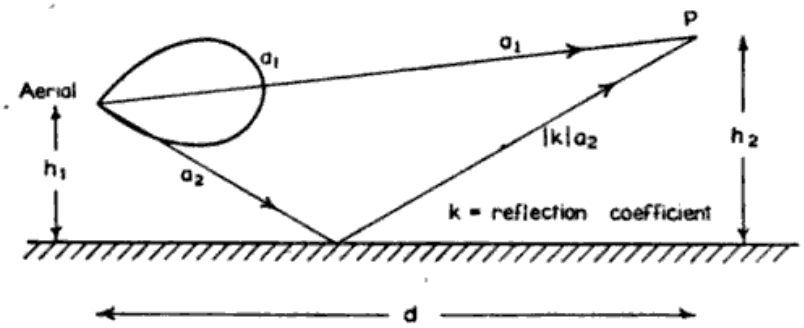


Figure 7.—Direct and reflected rays giving rise to interference phenomena above the horizon.

As the point P is raised the path difference increases and the field goes through a succession of interference maxima and minima. If the relative amplitudes a_1 and a_2 of the rays leaving the aerial are equal, then the maxima are of magnitude $1+|k|$ and the minima $1-|k|$ relative to the free space fields where k is the reflection coefficient. In the case of

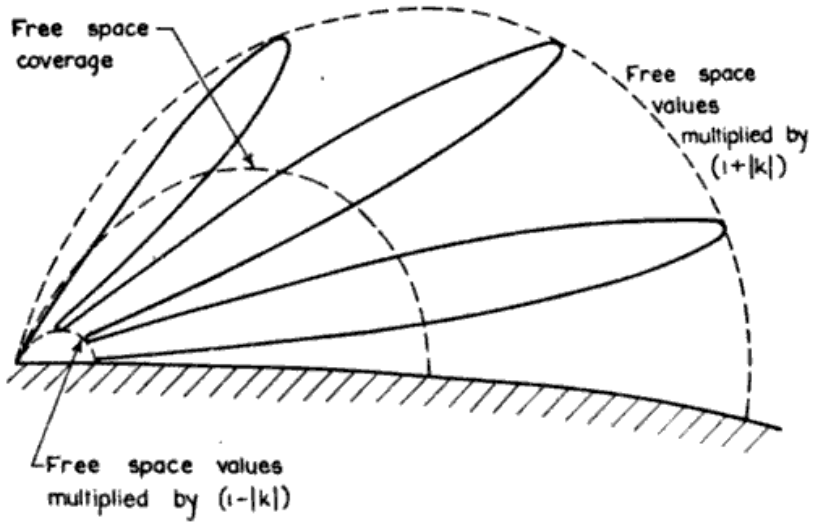


Figure 8.—Normal coverage diagram in presence of the earth.

the sea k is very nearly equal to -1 for both vertical and horizontal polarisation at small angles of elevation. The vertical coverage diagram is therefore modified in the way indicated in Fig. 8, a series of lobes being formed with the first minimum on the surface. The range of detection in the

middle of the lobes is almost doubled but at the expense of gaps in coverage.

Referring again to Fig. 7, if atmospheric refraction and earth curvature are neglected, the path difference Δ between the direct and reflected rays for h_1 and h_2 much less than d, is given by

$$\Delta = \frac{2h_1h_2}{d}.$$

For a reflection coefficient of -1, the resulting field strength for one-way transmission is modified by the factor $2 \sin 2\pi h_1 h_2/\lambda d$ which has a zero at the surface, and for low heights increases linearly with height. In this region, writing the angle for its sine, the modified form of the basic radar equation (4) becomes

$$d_o^4 = \frac{P}{p_o}\left(\frac{4\pi h_1 h_2}{\lambda d_o}\right)^4 \frac{A^2\sigma}{4\pi\lambda^2}$$

or

$$d_o^8 = \frac{P}{p_o}\frac{(4\pi)^3\, h_1^4\, h_2^4 \cdot A^2\sigma}{\lambda^6}.$$

It is of interest that d_o is proportional to $(P/p_o)^{\frac{1}{8}}$, $h_1^{\frac{1}{2}}$, $h_2^{\frac{1}{2}}$ and to $\lambda^{-\frac{1}{2}}$. For constant aerial gain, equation (5) applies and d_o is proportional to $\lambda^{-\frac{1}{2}}$. It is not permissible to neglect earth curvature except at rather short range, so that these equations are restricted in application. The theory for a spherical earth has been discussed by Jaeger.[3] It is found that as for a flat earth, the range of detection at low heights is greatly enhanced by the use of microwaves and by increasing aerial height.

The above treatment is applicable to propagation over the sea or smooth ground. Propagation over irregular ground is very complex.

Normal Refraction

So far we have neglected effects due to refraction. These occur even in a normal atmosphere as the refractive index decreases with height and so tends to bend rays downward.

[3] Jaeger's work is unpublished. See *Radio Wave Propagation* (N.D.R.C. Report), Academic Press, New York, 1949, pp. 53-63.

If the gradient of refractive index with height is uniform, its effect can be treated elegantly by a method due to Schelling, Burrows and Ferrell,[4] in which the coordinate system is transformed so that rays travel rectilinearly and the effective radius

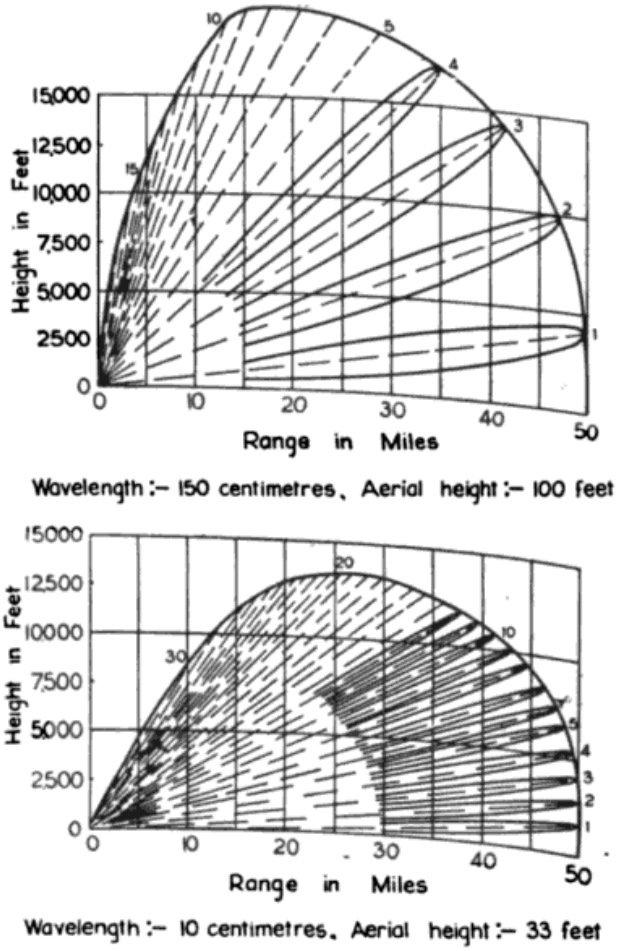


Figure 9.—Typical shipboard radar coverage diagrams.

of the earth is modified. This is usually a good approximation using an effective radius increased in the ratio 4/3. There is evidence that for low heights over the sea, especially in warm climates, a factor of 2/1 gives better results.

In Fig. 9 are shown vertical sections of typical shipboard radar coverage diagrams on 1·5 metres and 10 centimetres

[4] J. C. Schelling, C. R. Burrows and E. B. Ferrell, "Ultra-short wave propagation," *Proc. I.R.E.*, Vol. 21, March, 1933, pp. 427-63.

from which are apparent the outstanding features of increased number of lobes and decreased height of the bottom lobe with reduced wavelength. Land-based air warning equipments employ larger aerials, and ranges of detection of aircraft from 70 to 200 miles are usual.

5. Superrefraction

We have already mentioned that the normal atmosphere, because of its greater density near the ground, causes some bending of rays around the earth. This causes an increase of signal at points near to and beyond the horizon. It is quite common for this effect to be accentuated markedly by the presence of low lying masses of optically dense air. Under these conditions, ships, low flying aircraft, and land are often seen at distances up to one or two hundred miles by radar sets which have normal ranges of perhaps 20 miles. The phenomenon is known as superrefraction. The converse condition in which signals from low lying objects are reduced occurs but is less common.

Superrefraction has been studied extensively and a theory due to Booker[5] gives a good qualitative explanation of the phenomena. Such confirmatory measurements as have been made do not show good quantitative agreement, however, and it is not clear whether this is due to the paucity and complexity of the data or to the omission of an essential factor such as heterogeneity of the atmosphere.

The refractive index of air for most radar wavelengths differs from that for light in that water vapour is much more refracting in the former case. The collapse of dielectric constant of water from 81 at long wavelengths to about 1·7 at light wavelengths does not commence till near the absorption band at 1·2 centimetres. Consequently humidity plays a very important part in superrefraction.

[5] H. G. Booker and W. Walkinshaw, "The mode theory of tropospheric refraction and its relation to waveguides and diffraction," Physical and Royal Meteorological Societies' Report, 1947, pp. 80-127.

The index of refraction of moist air n is given by the semi-empirical formula

$$n = 1 + \left\{ \frac{79P_a}{T} + \frac{380{,}000e}{T^2} \right\} 10^{-6}$$

where P_a = partial pressure of dry air (millibars), e = partial pressure of water vapour (millibars), T = temperature (degrees Kelvin). We shall tentatively use ray theory and note that bending of rays is caused by gradients of refractive index. Quantitatively, for nearly horizontal rays the curvature ψ of the ray can be shown to equal the vertical gradient of refractive

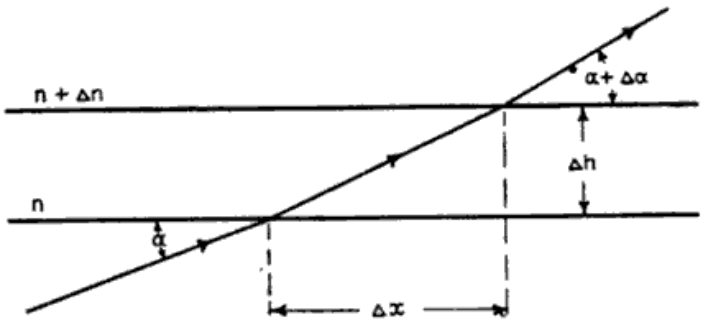


Figure 10.—Passage of a ray through horizontally stratified dielectric.

index as follows. Referring to Fig. 10 which depicts a ray passing through a horizontally stratified medium, Snell's law gives

$$\frac{\cos(\alpha + \Delta\alpha)}{\cos \alpha} = \frac{n}{n + \Delta n}.$$

For small values of α and n approximately unity this gives

$$1 - \Delta n \doteqdot \frac{1 - \frac{1}{2}(\alpha + \Delta\alpha)^2}{1 - \frac{1}{2}\alpha^2} \doteqdot 1 - \alpha\Delta\alpha .$$

Substituting $\Delta h/\Delta x$ for α and rearranging gives

$$\frac{\Delta n}{\Delta h} = \frac{\Delta\alpha}{\Delta x} = \psi$$

as stated in words above.

If we transform coordinates so that the earth is flat a ray will have a net curvature ψ_1, away from the transformed earth, given by

$$\psi_1 = \frac{dn}{dh} + \frac{1}{a}$$

where a is the radius of the earth. This transformation is illustrated in Fig. 11. We could, from measurements of

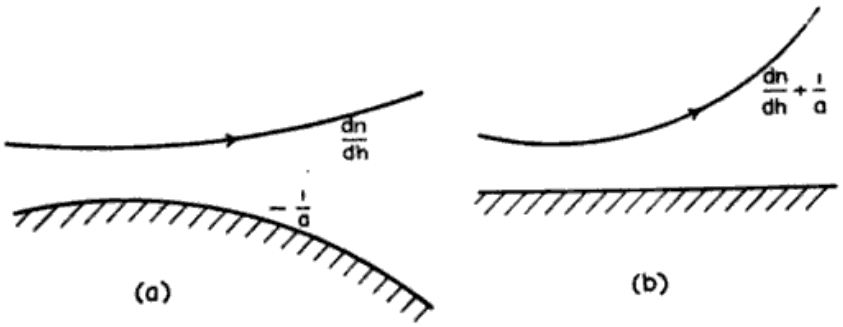


Figure 11.—Transformation to a flat earth. The symbols indicate curvatures away from the earth.

temperature and humidity over a range of heights, calculate n for each height, deduce dn/dh, and by adding $1/a$ determine whether a ray was being bent towards or away from the earth. The standard procedure now adopted is to introduce a new quantity, the "modified refractive index" M, defined by

$$M = \left[(n - 1) + \frac{h}{a}\right] 10^6,$$

where h is the height above an arbitrary datum. Modified refractive index is plotted against height. The slope dM/dh of this line then defines the relative ray curvature ψ_1, since

$$\frac{dM}{dh} = \left(\frac{dn}{dh} + \frac{1}{a}\right) 10^6 = 10^6 \times \psi_1.$$

In the formula defining M, the 1 and the 10^6 are introduced in order to give M values convenient for plotting (M is in typical cases about 300). In order to clarify the effect of positive and negative gradients of M, Fig. 12 illustrates ray paths for

the two simple cases of constant M gradient, (*a*) positive, (*b*) negative. If mean values of the variation of temperature and humidity with height are taken, as in the specification of the NACA[6] standard atmosphere, it is found that M varies approximately linearly with height, as in Fig. 12 (*a*), up to a considerable height. The gradient dM/dh is about + 35 per 1000 feet and the effective earth's radius is increased to 4/3

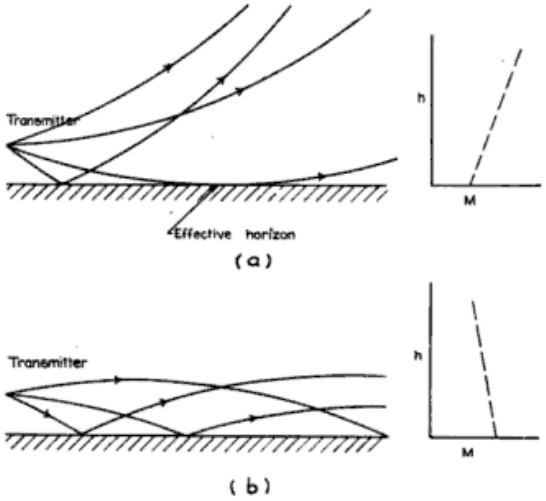


Figure 12.—Ray paths and M-h curves for constant dM/dh. (*a*) dM/dh positive. (*b*) dM/dh negative.

of its value neglecting refraction. As in the case without refraction there is a definite tangent ray and no ray can, on this restricted theory, penetrate into the shadow region beyond the radio horizon.

The case of Fig. 12 (*b*) is in marked contrast in that rays are being continually bent towards the earth and no shadow is

[6] W. S. Diehl, "Standard atmosphere—tables and data," National Advisory Committee for Aeronautics, Technical Report No. 218, 1925.

formed. This negative gradient of M cannot persist to great heights because the refractive index at the surface only exceeds unity by a small amount and cannot decrease below unity. Nevertheless it is only necessary for the radar set to be within the region of negative gradient for nearly horizontal rays to be prevented from escaping. This is illustrated in Fig. 13.

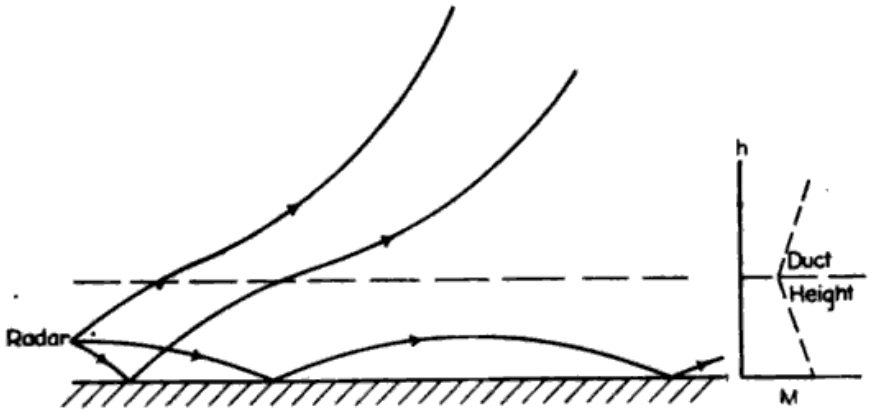


Figure 13.—Ray paths and M-h curve for simple duct with base on surface.

This phenomenon, which does not depend on the precise shape of the M curve but only on the location of the transmitter within a region of negative M gradient, is known as trapping. The region of negative gradient is known as a duct. A duct acts as a waveguide in guiding energy along its length from a source within it.

We have however omitted an important factor in treating the phenomenon from the point of view of geometrical optics. The theory is invalid if the structure is not large with respect to the wavelength. In this case the wavelength concerned is the vertical distance between successive points of equal phase, $\lambda \operatorname{cosec} \alpha$, where α is the angle of elevation. For nearly horizontal rays this can be large and in consequence situations showing negative M gradients will only produce substantial trapping if they extend over an adequate height for the wavelength concerned. The required value depends on detailed structure, but as a guide, 50 feet at a wavelength of 10 centi-

metres, and 750 feet at 150 centimetres are effective. Ducts which are not thick enough for the wavelength behave as leaky waveguides.

The case considered so far is that of a duct on the surface. Under certain conditions, the shape of the M curve may be complex giving rise to an elevated duct as illustrated in Fig. 14.

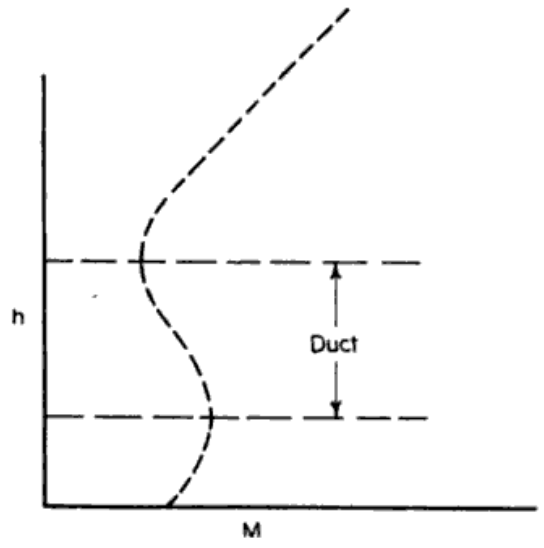


Figure 14.—M-h curve for elevated duct.

A great many measurements have been made of humidity and temperature of the air in the first few thousand feet above ground using aircraft, kites and tethered balloons and attempts have been made to relate these to radio measurements. In general it has been shown that the existence of pronounced superrefraction is associated with regions of negative dM/dh and there is good qualitative agreement between theory and experiment. On a quantitative basis there still remains some doubt but the subject is made very difficult by the usual complexity of a meteorological phenomenon. Normal conditions are illustrated in Fig. 15 (a), and conditions associated

with moderate superrefraction in Fig. 15 (*b*). The modification to a vertical coverage diagram which results from superrefraction is shown diagrammatically in Fig. 16.

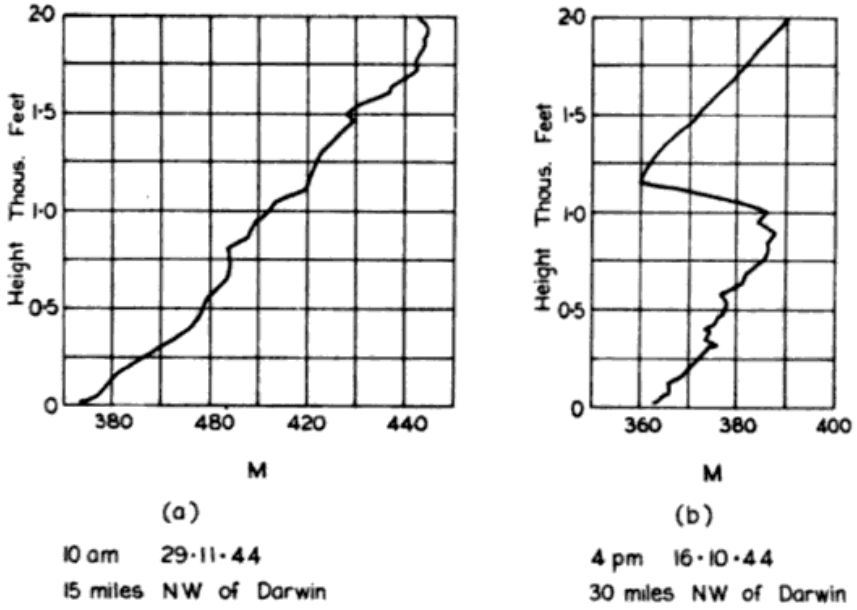


Figure 15.—Experimental M-h curves over sea near Darwin. (*a*) Typical normal atmosphere associated with onset of north-west monsoon. (*b*) Example of elevated duct (causing superrefraction) associated with sea breeze.

[*From C.S.I.R. Radiophysics Laboratory, Report RP* 260, 1945.

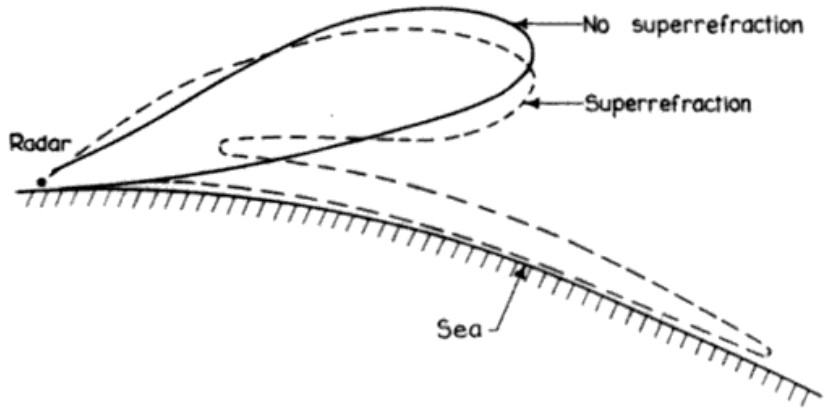


Figure 16.—Modification during superrefraction to bottom lobe of vertical coverage diagram of 1·5 metre radar due to surface duct of moderate strength. Higher lobes are not seriously modified.

Superrefraction occurs frequently in many parts of the world. It is caused when meteorological conditions are such that relatively cold or damp air underlies warm dry air. The

situation which has received most attention following the pioneer work of Booker and Woodward[7] is the off-shore streaming of relatively hot dry air over the sea. Evaporation takes place and if the air was originally warmer than the sea it cools and is convectively stable, so that a blanket of cool humid air is formed at the surface. This condition is most frequently developed when trade winds blow from a desert over the ocean, and spectacular examples are known to occur off the coasts of the Sahara Desert, Arabia and north-west Australia.

A second important coastal mechanism giving rise to superrefraction in temperate latitudes is that known as "the coastal front" by Australian meteorologists. It is essentially the sea breeze mechanism (bringing in sea air under the continental air) engendered by a considerable contrast in temperature between land and sea. When well developed off a continental coast the effect can persist throughout both day and night. It occurs when the motion of the continental air mass brings hot dry air to the coast so that the general meteorological situation is similar to that for off-shore streaming described above, but differing in that the local winds are on shore and not off shore. This mechanism is predominant on temperate coasts, as for example off the southern half of Australia in summer. Both these mechanisms show a diurnal variation with an afternoon and evening maximum.

Over land, superrefraction can be as pronounced as at sea. It occurs at night and is due to the cooling of the land and lower levels of the atmosphere by radiation. It occurs principally during fine cloudless nights and the detailed mechanism is complicated by evaporation from or deposition of dew on the ground and has not been extensively studied. It is of some interest that such a radiation inversion can cause superrefraction over the sea as it blows outward.

[7] H. G. Booker, "Elements of radar meteorology: How weather and climate cause unorthodox radar vision beyond the geometrical horizon," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 69-78.

These three mechanisms tend to give ducts of the order of 1,000 feet in height and tend to influence radar propagation similarly in the range of wavelengths from 3 to 150 centimetres.

Over the sub-tropical oceans of the world in the region of the trade winds a further mechanism is operative. The air is blowing into regions receiving increasing amounts of solar energy. This energy is chiefly effective in evaporating water. Consequently a layer of humid air is formed over the ocean. When the wind is weak the layer of excess humidity is very shallow; when strong, increased turbulence increases the thickness of the layer. The height of the duct so formed is roughly proportional to wind speed and is about 40 feet for a 10 knot breeze. These ducts cause pronounced super-refraction but only at short wavelengths. Radar sets on 3 and 10 centimetres show marked effects more particularly if the aerial is low so that it lies in the duct.

Under normal atmospheric conditions echoes from ships or land are seen on powerful surface radar equipments up to distances of 20 to 40 miles. Superrefraction which increases these distances to 100 to 200 miles is common, occurring near warm temperate coasts perhaps one tenth of the total time. Less frequently and only in favourable localities much more spectacular increases are observed. Thus 1·5 metre air warning sets near Darwin, Australia, report echoes from the coast of Timor, 300 to 500 miles distant, several times a month.[8] A similar equipment near Broome in north-west Australia reported echoes in November, 1944, from the coast of Java 900 to 1,100 miles away.

Intense superrefraction also occurs in the Arabian Sea. Echoes on 1·5 metres from a number of places on the coast

[8] F. J. Kerr, "Radio superrefraction in the coastal regions of Australia," *Aust. J. Sci. Res.* A, Vol. 1, 1948, pp. 443-63.

from Aden to Karachi have been obtained from Bombay at distances up to about 1,500 miles.[9]

Ranges of detection of ships which are normally of the order of 20 miles are, similarly, frequently increased to 100 miles. It is unusual for a range greater than 200 miles to be reported but an extreme range of 700 miles has been reported from Bombay.

Normal ranges of detection of aircraft at medium heights are not so severely limited by the shadowing effect of the earth as those of surface targets, and range increases due to superrefraction are relatively less. The usual effect is an increase of range on low flying aircraft to 100 or 200 miles and it is suspected, but not proved, that this is associated with some reduction of range to high flying aircraft. Occasional increases of range are reported greatly beyond the normal even for high flying aircraft. Thus in February, 1944, the air warning station at Geraldton, West Australia, followed a Catalina flying boat, outward bound from Perth to Colombo, almost continuously out to a distance of 800 miles.[10]

It is noteworthy that all the above outstanding increases of range of detection have been obtained on 1·5 metres, but it is not clear whether this should be attributed to specially favourable propagation or to the particular distribution of radar equipments from which reports are available.

6. Atmospheric Absorption

Absorption in air at ordinary radio frequencies is negligible but at the highest frequencies used in radar, this is not so.[11]

[9] H. G. Booker, "Elements of radar meteorology: How weather and climate cause unorthodox radar vision beyond the geometrical horizon," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 69-78.

[10] "Abnormal ranges on aircraft in western area," R.A.A.F. Report to Radiophysics Laboratory, March, 1944.

[11] J. H. Van Vleck, "The absorption of microwaves by (*a*) oxygen and (*b*) uncondensed water vapour" (two papers), *Phys. Rev.*, Vol. 71, 1947, (*a*) p. 413, (*b*) p. 425.

Two atmospheric gases, oxygen and water vapour, show absorption which may be regarded as due to extreme infra red absorption bands associated with the magnetic dipole moment of oxygen and the electric moment of water. Fig. 17 gives

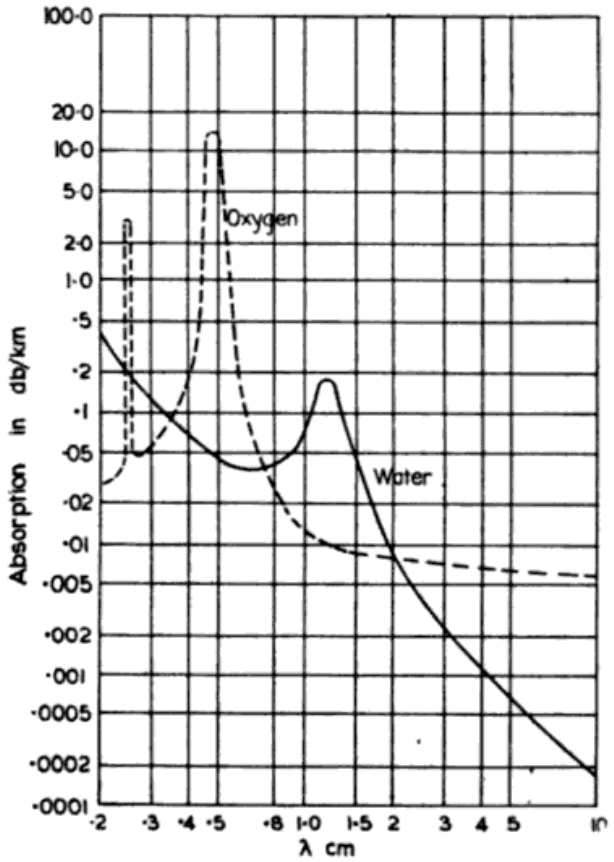


Figure 17.—Absorption in atmospheric gases (one way attenuation). (*a*) Water vapour of partial pressure 7·6 millimetres of mercury. (*b*) Oxygen in a normal atmosphere of total pressure 76 centimetres of mercury.

theoretical attenuation figures for a typical atmosphere. The attenuations are proportional to the partial pressures of the respective gases and are not rapidly varying with temperature.

The attenuation will be highest in the tropics and may reach about 0·1 decibels per kilometre at a wavelength of 3 centimetres and 0·5 decibels at 1·2 centimetres, the centre of the water vapour absorption band.

The atmosphere may also carry solid or liquid particles in

the form of fog, cloud, rain etc. and, as in the case of light, these particles cause scattering and absorption. The absorption is trivial in the case of the longer radar wavelengths but is large in the one centimetre region. Fig. 18 gives estimated

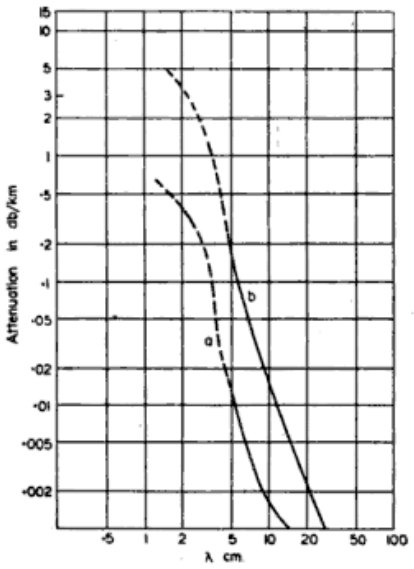


Figure 18.—Attenuation, one way, due to rain. (*a*) Moderate rain (6 millimetres per hour of known drop size distribution). (*b*) Rain of cloud burst proportion (43 millimetres per hour of known drop size distribution).

values for moderate and extremely heavy rain. Values for cloud are much less. A rough generalisation is that attenuation due to rain becomes appreciable at a wavelength of about 10 centimetres, that due to cloud at 1 to 3 centimetres.

For drops which are small with respect to the wavelength the absorption is proportional to the total water content per unit volume.

7. Atmospheric Scattering

In a manner analogous to that in which light is scattered by fine dust particles, radar waves are scattered by rain drops.[12] The amount of scattering from heavy rain is adequate to produce echoes on powerful 1·5 metre sets and is very large on microwave equipments. Excepting for very large drops at the shortest wavelengths, the echo power reflected from a single drop obeys Rayleigh's law, being proportional to D^6/λ^4 where D is the diameter. Consequently for uniform rain of density N drops per unit volume, the total energy returned is proportional to ND^6 which may be considered as the product of total mass of water per unit volume and mass of a single drop. It is therefore sensitive to drop size whereas attenuation is not.

If on the other hand we vary wavelength the echo intensity does not usually show a simple dependence on $1/\lambda^4$ because we usually simultaneously vary a number of other controlling factors, such as aerial gain, pulse length, etc.

Because echoes are adequately intense and absorption not excessive, microwave radar equipments can be used to observe the structure of rain clouds in a way which is not possible by eye owing to the heavy attenuation of light. Plate II (*a*) shows the distribution of heavy rain echoes obtained during a typhoon. It is a photograph of the PPI display of a medium power 10 centimetre radar set on a ship located about 40 miles from the "eye" of the typhoon which is clearly visible.

Plate II (*b*) shows not a plan view but an elevation of a less spectacular cloud formation obtained on a 3 centimetre equipment with an RHI (range-height-indication) display.

[12] J. W. Ryde, "The attenuation of centimetre radio waves and the echo intensities resulting from atmospheric phenomena," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 101-3.

CHAPTER III

THE MAGNETRON

1. Early Magnetrons

For many years prior to 1946 the magnetron was studied and used as a generator of high frequency oscillations. Its construction followed the well known lines of a cylindrical anode surrounding a filamentary cathode, and it operated in an axial magnetic field. In many cases the anode was split into two or four segments, alternate segments being connected together in the latter case.

Such devices were used to excite oscillatory circuits of the Lecher wire type connected between filament and anode, or between anode segments, and power was generated at wavelengths between one metre and one centimetre approximately.

It was known that there was more than one mechanism whereby oscillations were produced, and most interest centred on the generation of power at the shortest wavelengths where the principle of operation was least understood. Much ingenuity was used and there were many variations in design; but it cannot be said that startling results were forthcoming in the range of wavelengths around 10 centimetres in which we are at present interested.[1]

Some typical results obtained at the shortest wavelengths will now be noted. Rice,[2] using a single anode magnetron, produced 3 watts at a wavelength of 4·8 centimetres, the efficiency being only 1 per cent. Cleeton and Williams[3] produced very low powers at wavelengths of 1·87, 1·22, and

[1] A useful review of early work, with a bibliography, is given by A. F. Harvey, *High Frequency Thermionic Tubes*, Chapman and Hall, London, 1943.

[2] C. W. Rice, "Transmission and reception of centimetre radio waves," *Gen. Elec. Rev.*, Vol. 39, 1936, pp. 363-68.

[3] C. E. Cleeton and N. H. Williams, "The shortest continuous radio waves," *Physical Rev.*, Vol. 50, November, 1936, p. 1091.

0·64 centimetres. Linder[4] with two different designs working at about 10 centimetres, produced powers of 2·5 watts (12 per cent efficiency), and 13 watts (20 per cent efficiency).

In view of developments to be described we shall not dwell on pre-war work on the magnetron except to glance at the design used by Rice, which is shown diagrammatically in

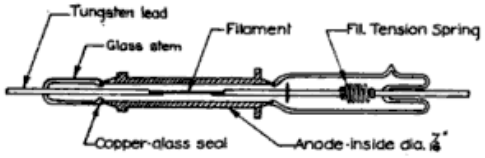


Figure 19.—Single anode magnetron—as used by Rice (1936).

Fig. 19, and the anode used by Linder, Fig. 20. The Rice magnetron was designed so that the anode formed part of a coaxial transmission line resonator, while the anode of the Linder magnetron was split so as to form part of a two-wire transmission line resonator. These were both attempts to

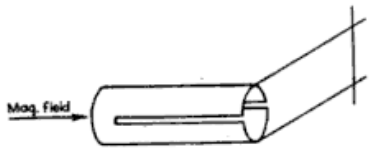


Figure 20.—Magnetron anode system—as used by Linder (1939).

resolve the conflict between high power which requires large dimensions, and short wavelength which requires the opposite. We will now see how this problem was finally solved.

[4] E. G. Linder, "Description and characteristics of the end-plate magnetron," *Proc. I.R.E.*, Vol. 24, Part I, 1936, p. 633-53; and "The anode-tank-circuit magnetron," *Proc. I.R.E.*, Vol. 27, 1939, pp. 732-38.

2. The Resonator Magnetron

The modern high-power magnetron was developed in the Physics Department of the University of Birmingham, with the co-operation of the General Electric Company (Wembley), as part of a nationally inspired programme of work on wireless waves of a frequency of 3000 megacycles per second and greater. We shall briefly describe this very remarkable achievement, and in doing so we cannot do better than quote from Oliphant's foreword to the report which describes the original work.[5]

"An examination of the very recent literature and of first principles showed that the only hope for producing efficient generators was to combine generator and circuit in a single unit, and that the circuit should be one of high efficiency made from the best possible electrical conductors. Accordingly, a programme was drawn up for detailed investigation of the velocity modulation methods which had just been published using efficient 'rhumbatron' resonators, while a paper by Rice in the General Electric Review for 1936 which describes a transmission line type of magnetron, encouraged us to investigate the possibility of introducing rhumbatron technique to improve the circuit of the magnetron. A very superficial examination of existing magnetron devices showed that circuits were made of highly resistant materials, and in a form where radiative and resistive damping seriously reduced the efficiency. Plans were drawn up of possible trial apparatus and the work was entrusted to Dr. Randall and Mr. Boot. They found it was not at all easy to transform the existing types of resonator rhumbatrons for use in the magnetron, where a cylindrical symmetry was desirable, and with considerable insight decided to try the less efficient cylindrical form of resonator, which was at once successful. The general ideas formed at that time with regard to mode of action and general construction have been a little modified by subsequent work, though what was rather a crude laboratory instrument

[5] J. T. Randall, H. A. H. Boot and S. M. Duke, "Magnetron development in the University of Birmingham," Co-ordination of Valve Development Report, Department of Scientific Research and Experiment, Admiralty, 1941.

was transformed through the experience of the Research Laboratories of the General Electric Company into a valve of great simplicity which was easily manufactured and was reliable enough for service use."

The first valve produced is shown in Plate III (*a*). The anode system with its six cylindrical resonators was machined out of a solid copper block, the block itself with metal end-plates forming the main valve envelope. The resonator size was determined on the basis of the Hertzian dipole ring for which $\lambda = 7{\cdot}94d$, where d is the diameter of the ring. Power was taken out through a coaxial seal terminating in a coupling loop in one of the resonators. The cathode was a pure tungsten filament 0·75 millimetres in diameter. This arrangement provided an immediate solution to the pressing high-power magnetron problem. It introduced high Q resonant circuits of the right size and permitted high anode dissipation. Glass bombardment by electrons escaping from the ends of the anode did not exist.

This magnetron, running on the pumps, worked for the first time on February 21st, 1940. Approximately half a kilowatt of power at 10 centimetres was generated, the power input being 4 kilowatts. This was a tremendous advance over the performance of any previous magnetron. The anode voltage was 8,000 and the magnetic field 0·11 webers per square metre (1100 gauss). The anode block dimensions were : resonator diameter 1·2 centimetres, anode diameter 1·2 centimetres, slot width 0·1 centimetres, and slot length 0·1 centimetres.

The next step was the construction of sealed-off tubes, which was made possible by the use of the General Electric Company gold-ring seal to be described later. An indirectly heated, oxide coated cathode was also added and this made a material improvement in efficiency. Developments were rapid and large numbers of magnetrons of various operating frequencies and powers were made, and used by the Armed Services in subsequent years. Production and development of these tubes was carried out in laboratories and factories throughout the English-speaking world, including Australia.

Information from other countries is at present meagre, but

there is evidence of advanced Russian work in this field.[6] It is also known that much work as yet unreported in detail was done in Japan and Germany. Magnetrons designed in those countries used resonator systems somewhat similar to the one described above, but mounted in large glass envelopes. They performed badly, and were difficult to construct.[7]

3. Theory of Magnetron Operation

Before describing further developments, let us see if we can form some picture of how these magnetrons work. Following a useful discussion by Slater[8] we shall consider electron motions in the simple case of a plane magnetron with suitably arranged electric and magnetic fields. Space charge effects will be ignored and initial electron velocity will be taken as zero.

Referring to the conditions of Fig. 21, simple analysis

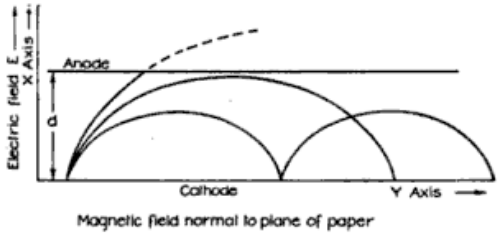


Figure 21.—Electron paths in steady crossed electric and magnetic fields.

shows that the electron paths are cycloids traced out by a point on a circle of radius $R = mE/eB^2$ rolling along with angular velocity $= Be/m$ and with linear speed E/B, where e and m

[6] N. F. Alekseev and D. D. Malairov, "Generation of high-power oscillation with a magnetron in the centimetre band," *Proc. I.R.E.*, Vol. 32, March, 1944, pp. 136-39.

[7] A study of the literature, including patents, shows that some of the ideas embodied in the Birmingham magnetron, including the association of cavity resonators with a number of segments, had occurred to previous workers at one time or another. Credit must be given to the Birmingham group, however, for the development and construction of the first practical high-power magnetron from which has sprung a whole series of sturdy valves operating over wavelengths from 1 to 50 centimetres, and generating peak powers up to several megawatts.

[8] J. C. Slater, "Theory of the magnetron oscillator," Massachusetts Institute of Technology Radiation Laboratory, Report, August, 1941; J. B. Fisk, H. D. Hagstrum, P. L. Hartman, "The magnetron as a generator of centimetre waves," *Bell Syst. Tech. J.*, Vol. 25, 1946, pp. 167-348.

are the charge and mass of the electron, E is the electric field strength and B is the magnetic flux density. Three paths are indicated for different values of E/B^2.

It is noteworthy that at the cusps near the cathode plane the kinetic energy of the electron is zero, whereas at the top of the path the work done by the field on the electron appears as kinetic energy. It is also useful to remember that with respect to axes moving along at speed E/B, the coordinates of the electron's position at any time can be written

$$X' = R\cos\omega t, \qquad Y' = R\sin\omega t.$$

Suppose that there is superimposed on the steady electric field a harmonic electric field of a period equal to $2\pi/\omega$ the period of rotation of the electron. The electron paths can be computed by direct solution of the equations of motion, but all we require here is a qualitative idea of these paths which can be obtained by thinking of the effects of the sinusoidal field on the simple harmonic components of the electron's motion. When the field is in phase with those components, the rotational motion of the electron is built up as the field does work upon it, while a field opposite in phase will slow down the rotational motion. Picturing only the first few cycles of either case, the paths shown in Fig. 22 are obtained. The energy exchanges

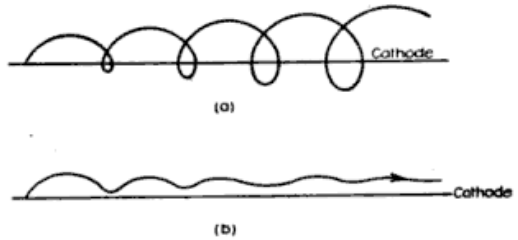


Figure 22.—Electron paths in crossed electric and magnetic fields, with added sinusoidal electric field.

are important: in the first case the electron absorbs energy from the alternating field, in the second the electron gives up energy to the alternating field. These are of course the extreme

cases ; electrons emitted at other times either receive, or give up energy more slowly. Thinking for a moment of the second case which is the more favourable for yielding energy to the alternating field, we see that the electron finishes up with its drift speed E/B only, its rotational motion having been lost by the action of the alternating field. It is moving at a distance R from the cathode. Hence substituting from above, the work done on it by the steady field is,

$$EeR = Ee \cdot \frac{mE}{eB^2} = m\frac{E^2}{B^2}$$

which equals twice the kinetic energy of drift $\frac{1}{2}m(E/B)^2$. Thus in a favourable case, the electron is able to give up to the alternating field one half of the energy extracted from the steady field.

There is one vital point, however, about paths of the first type which must not be overlooked. All such paths cross the cathode because of the increase of speed due to the gain of energy from the field. They will therefore strike the cathode after one period and have no further opportunity of absorbing energy.

From this simple case we can now understand why a single anode magnetron can be expected to oscillate. Remembering that electrons are emitted from the cathode in all phases, we see that there are two groups. Roughly half the electrons absorb energy from the alternating field, but are removed after one period and thus do not have much damping effect on the alternating-field generator. The remainder, which can yield up to the alternating field about half of the energy they have received from the steady field, will clearly tend to sustain oscillations once they commence.

It should be noted that the direction of the superimposed field in the plane of the orbits is not significant. It must not be forgotten that the " spent " electrons which have delivered energy to the field must be removed before they can absorb energy again by building up their rotational motion and that only oscillations of a frequency roughly equal to the natural electron frequency $\omega = Be/m$ have been considered.

A magnetron working in this way will generate power at a wavelength $\lambda = 2\pi c/\omega = 2\pi mc/eB$. We have therefore $\lambda B = 10{\cdot}6 \times 10^{-3}$ where λ is in metres and B in webers per square metre; hence if λ is in centimetres and B in gauss, $\lambda B = 10{,}600$.

This relation will be found to be approximately satisfied in the case of the simpler single and split-anode cylindrical magnetrons used by the workers mentioned earlier.

The Multi-segment Cylindrical Magnetron

Turning now to the electrode system used in the Birmingham magnetron, the system becomes cylindrical instead of plane and the electric field is complicated by the presence of the transverse fields across the slots. Exact tracing of the electron

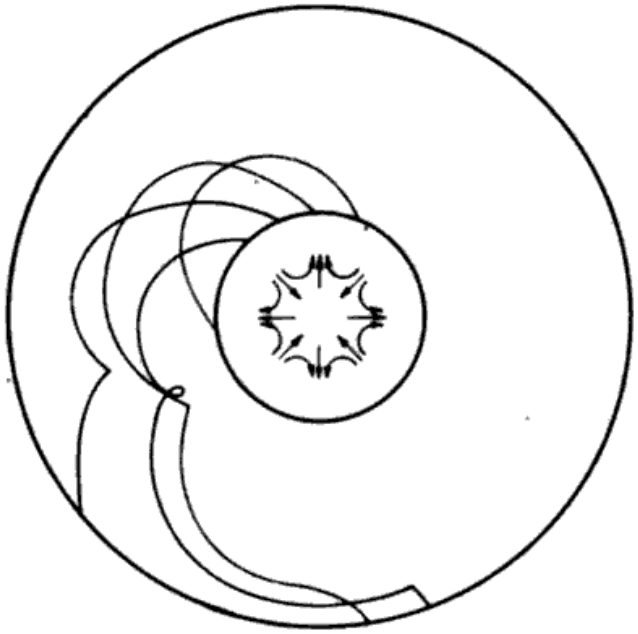


Figure 23.—Electron orbits in a magnetron with 8 segments. The wavelength is 10 centimetres, cathode radius 3 millimetres, anode radius 8 millimetres, anode potential 10 kilovolts, AC potential 5 kilovolts, magnetic flux density 0·13 webers per square metre (1300 gauss), assuming a linear potential. Arrows show the phase of the electric field at the cathode at the instant of emission.

orbits is difficult even if space charge be neglected, but it can be shown that the electrons move around the system in cycloid-like paths, and can both absorb energy from and deliver it

to the alternating fields which are produced across the slots by the resonators. When the drift speed of the electrons has the right relation to the period of the resonators so that electrons pass each slot in about the same phase, we find that the energy-

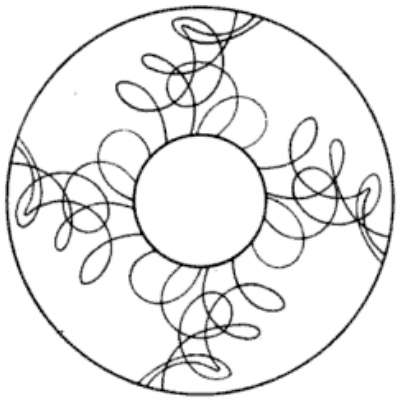


Figure 24.—Orbits of Fig. 23 as seen from rotating co-ordinates.

absorbing electrons are removed, while the energy-yielding electrons move out and strike the anode with very little residual energy. This represents an important change from the simple case, when the spent electrons remained in the middle of the discharge space to absorb energy. In the multi-segment magnetron no such problem arises, and some of its increased efficiency may be attributed to this fact.

Efficiency is also increased by the fact that the residual energy of some electron orbits is much below 50 per cent of the work done on them, making the "orbit efficiency" high. These are the orbits with cusps close to the anode. In the plane case maximum orbit efficiency is 50 per cent.

As a practical example, Figs. 23 and 24 show orbits for an eight-segment magnetron, calculated by Slater by numerical integration of the equation of motion. There has as yet been no mention of space charge effects, but what is believed to be a

fair approximation to take care of this has been made in computing the orbits shown (Slater,[8] p. 118).

The orbits are of familiar shape. Three of them have a chance of giving energy to the field; the fourth begins to absorb energy, but is lost before a complete cycle occurs. The energy calculations show that of the work done on these four electrons, only 60 per cent appears at the ends of the orbits, and 40 per cent is available for maintaining oscillation.

The grouping of orbits shown in Fig. 24 gives an interesting physical picture of the cloud of electrons rotating as a toothed wheel with the speed of the travelling voltage wave at the anode. This picture turns out to be useful in the understanding of space charge effects.

Much work[8,9] has been done on the theory of the multi-segment magnetron, investigation having been carried out on:

(*a*) The behaviour of the multi-cavity resonator system, its modes of oscillation and the fields it produces in the working space.

(*b*) Electron concentrations and paths in the working space under the influence of fields due to the resonators and space charge.

The problem has to be split up in this way to make headway in the face of its complexity and numerical integration must be resorted to at most stages.

A brief consideration of modes of oscillation is given in the next section and we will now consider problem (*b*). Here the greatest success has come from a "self-consistent field" method of approximation. In Slater's words, the problem is: "Given a certain boundary condition in the form of a field at the anode surface, what must be the electric field in the anode-cathode region, in order to produce the motion of the electrons which will in turn set up the space charge necessary to produce the field?" The method is to make a reasonable guess at a field and compute orbits and space charge. Then

[8] A series of reports of work done by teams at Manchester and Leeds Universities, led respectively by Dr. Hartree and E. C. Stoner, has been issued as Co-ordination of Valve Development Reports by the Admiralty Department of Scientific Research and Experiment, numbered Mag. 1, 2, 3 etc., commencing August, 1941.

by deriving a field from this space charge (with perhaps some smoothing), the orbits are re-computed and so on. Convergence of assumed and derived fields gives the "self-consistent field" solution. Examples of such a computation with a discussion of its problems are given by Hartree and others.[9] As a starting point, use is often made of the space charge disposition which has been employed with some success in the single-anode magnetron. It should be noted that a complete calculation will give figures for emission and anode currents as well as energy exchange.

While some success attended the efforts of those working on this difficult problem, Slater's opinion in 1941 was that there was at that time no completely satisfactory solution of the space charge problem. He pointed out however that whatever uncertainty there was in the region close to the cathode, the electron orbits were generally little changed by variation in the assumed space charge disposition. If we recall Fig. 24, we may say that we do not know all about the hub of the rotating toothed wheel but we are fairly certain that the orbits in the teeth approach reality, and this is where most of the energy exchanges take place. A good investigation of magnetron operation may therefore be made without a serious consideration of space charge at all, and this Slater has done. He has been able to specify a number of points for practical magnetron design. These are briefly noted in the next sub-section.

Magnetron Design

Because of the difficulty of the theory, most magnetron design has proceeded empirically guided only by general theoretical conclusions. Slater deduces that for efficient operation:

(*a*) The drift speed of the electron must be such that the electrons pass each slot in about the same phase.

(*b*) The magnetic field should be above "cut-off." This is of course to give the useful electrons an orbit with at least one cusp, with a good chance of giving up their energy.

(*c*) The distance between slots should be roughly the same as the anode-cathode distance. This is to allow a reasonable tangential AC field component all over the working space.

None of these conditions is at all critical and it is not surprising to find that a magnetron will work over a large range of voltages, fields, and currents.

Equations describing (*a*) and (*b*) above for a multi-segment cylindrical magnetron are:

$$V = \frac{300\pi}{n\lambda}(\gamma_a^2 - \gamma_c^2)\left\{B - \frac{0{\cdot}0106}{n\lambda}\right\} \times 10^6 \qquad (1)$$

$$V = 22{,}200\, B\, \frac{(\gamma_a - \gamma_c)^2(\gamma_a + \gamma_c)^2}{\gamma_a^2} \times 10^6 \qquad (2)$$

where V is the steady anode potential, B is the magnetic flux density, λ is the magnetron wavelength, γ_c and γ_a are the anode and cathode radii; and n is the "mode number" or number of repeats of the voltage around the anode.

Hartree,[10] also, produced from other considerations an equation differing slightly from (1). The "Hartree diagram," well known in the literature, shows the cut-off curve plotted from equation (2); and a series of straight lines plotted from his equation for various values of n.

These equations have proved very useful as guides to magnetron design. They only go part of the way, of course, because questions of cathode emission, anode current, anode and cathode heating, and resonator dimensions must also be considered if a magnetron is to be designed *ab initio*. They can be used for "scaling"—the means whereby, given a successful magnetron, designs for higher or lower frequency operation are produced.

It is an interesting fact that most practical magnetrons whose dimensions were chosen by considerations of resonator size and manufacturing technique, require such high voltages and currents that pulse operation is necessary to prevent overheating of anode and cathode.

[10] H. A. Boot and J. T. Randall, "The cavity magnetron," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, p. 935.

4. Modes of Oscillation

It did not take early users of multi-segment magnetrons long to discover that even the most carefully made valves could oscillate at more than one frequency. It is easy to see why this can happen if we reflect that there may be 8 resonant circuits with some degree of coupling between each, so that a number of modes of oscillation is expected on general grounds. A simple example will show the type of " moding " which can exist. Fig. 25 gives a view of the segments of an

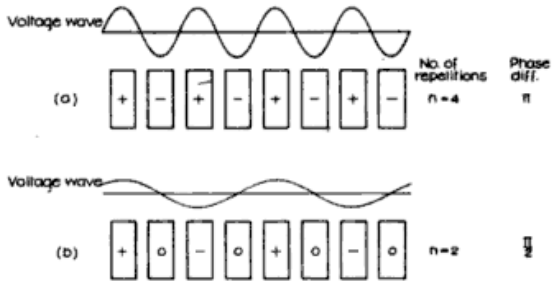


Figure 25.—Moding in an 8-segment magnetron. Segments shown " unwrapped," as seen from the cathode.

8 resonator magnetron which has been " unwrapped " as seen from the cathode. Two possible modes are shown, the charges on each segment being indicated by signs. Another way of describing these conditions is by giving the voltage wave for each case ; it may be considered as travelling in either direction, or as a stationary wave.

Mode (*a*) is described as the " π mode " or $n = 4$ mode, π being the phase difference between adjacent segments and n the number of repetitions of the wave pattern around the anode. The $n = 2$ mode is also shown and it is clear that there are modes of this type corresponding to $n = 1, 2, 3$ and 4 for the 8 segment magnetron. No further elaboration of this complex question will be made here. An important point,

however, is that even with perfectly made anodes, these modes give rise to different wavelengths which unfortunately are closely grouped—hence the difficulty in operation.

Strapping

The above analysis indicated to early workers with the Birmingham magnetron that some measure of mode selection could be attained by means of physical connection between segments. For instance if alternate segments of Fig. 25 are strapped electrically, the $n = 4$ mode would appear to be favoured at the expense of mode $n = 2$. These ideas were tested and found to be along the right lines, by "cold block" experiments in which the resonant behaviour of the anode

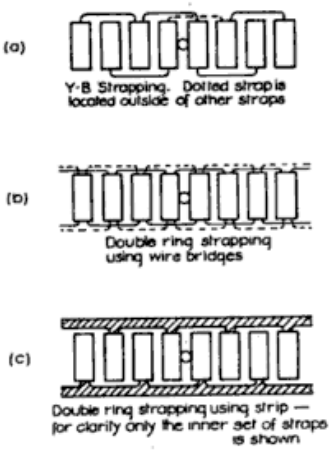


Figure 26.—Methods of strapping. Segments are shown "unwrapped" and the position of the coupling loop is indicated by a small circle.

block was examined by injection of a steady signal. Strapped magnetrons proved to be very free from "moding" troubles, i.e. it was difficult to produce energy at more than one fre-

quency and the efficiency of operation was higher. A slight drawback was that strapping increased the change of frequency with load impedance.

Fig. 26 shows strapping methods which have been used in practice, unwrapped magnetron segments being shown in each case and the position of the output coupling loop indicated by a small circle. Strapping is by wire or strip bridges mounted on the ends of the anode segments. Magnetrons strapped according to these methods have been shown to operate with high efficiency in the π mode (or $n = 4$ for eight-hole blocks).

The exact effects of strapping have been investigated thoroughly by Slater[11] and it appears that the presence of straps produces frequency separation of modes and a distortion of the field pattern of unwanted modes, which tends to make operation in the unwanted mode difficult. One point which has arisen and been found useful in practice is that a break in the strapping to give "incomplete strapping" (of which Y-B strapping is an example) helps in suppressing modes near the one desired. Incomplete strapping is always arranged symmetrically with respect to either the output loop or the cathode support. Such breaks exist in the strapping of magnetrons to be described later.

An important practical point is that strapping permits some degree of frequency adjustment during assembly. The block with straps assembled, is fed with a signal of the desired frequency and the whole tuned to resonance by bending the straps. This feature is much appreciated by manufacturers who have a rigid frequency tolerance to satisfy.

5. General Characteristics

Operating Conditions: Characteristic Diagrams

As there is no theoretical guide to the precise operating conditions of a magnetron, a series of measurements is usually made and presented in the form of a characteristic diagram. The coordinates are anode voltage and anode current, and

[11] J. C. Slater, "Resonant modes of the magnetron," Massachusetts Institute of Technology Radiation Laboratory, Report 43-9, August, 1942.

contours of constant power output, efficiency, frequency, and magnetic field may be drawn. An example is shown in Fig. 27. On this diagram peak readings of volts, amperes and kilowatts are given, as this characteristic has been taken under the pulse conditions which we have seen are necessary for high power magnetrons.

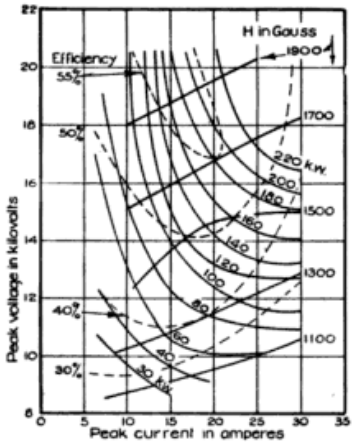


Figure 27.—Magnetron characteristics showing the relation between peak voltage and peak current for various magnetic fields. Contours of constant power output and efficiency are shown.

The characteristic diagram gives a good picture of magnetron behaviour, and enables suitable operating conditions to be fixed. It also shows the DC resistance offered to the source of pulse voltage, which information is often required for pulse network design. Discontinuities in any of the contours are matters of importance as they expose regions of instability which may result in "double moding." The example given is free from such defects.

Frequency Stability

Although the frequency of these magnetrons is nominally fixed, we find that there is nearly always a smooth change of frequency with magnetic field and peak current, as one would expect from space charge movements and the associated variations of effective dielectric constant in the working space. Such changes are usually exposed by the characteristic diagram, and with proper control of operating conditions need not cause trouble except in connection with the RF spectrum referred to below.

Change of frequency with block temperature is almost inevitable and in practice imposes a warm-up period during which tuning adjustments may have to be made.

RF Spectrum

The energy emitted by a magnetron extends over a range of frequencies, and the most desirable state of affairs exists when it is concentrated in one narrow band. Excitation of more than one mode, which unfortunately is always possible due to some peculiarity in design, manufacture, or operation, results in a serious loss of useful RF energy in most radar systems. With only one mode excited, RF bandwidth depends both on the magnetron itself and on pulse shape and duration. For further discussion of this important point see, for example, Collins.[12]

Effect of Load Changes on Frequency and Power Output.

Another operating difficulty may now be referred to. Particularly with magnetrons delivering full power it is found that the frequency of oscillation is affected to a disturbing degree by changes in the impedance presented by the feeder. Such behaviour is not surprising when we consider that the internal resonators are closely coupled to the external circuit.

[12] G. B. Collins (Ed.), *Microwave Magnetrons* (Massachusetts Institute of Technology, Radiation Laboratory Series, Vol. 6), McGraw-Hill, New York, 1948, pp. 54-6, 66-9, 83-92, 135-40.

Fortunately, it is possible to minimise these effects by a proper choice of mean load impedance, and an excellent method of achieving this has been put forward by Collins,[13] who has made a systematic experimental study of the loading of a number of types of magnetron. The magnetron under test was fed through a slotted section with a voltmeter probe as used in normal impedance measuring gear, into a transformer and load which were capable of presenting a large range of impedance to the valve. With a fixed anode current, measurements of power output, frequency and sometimes anode voltage were made for various impedances which were noted implicitly in terms of standing wave ratio and distance of the voltage minimum from some reference point.

These data were plotted on a Smith chart,[14] giving one form of what is now called a Rieke diagram. An example is given in Fig. 28. The polar coordinates (r, φ) are given by

$$r = |K| = (1 - S)/(1 + S)$$
$$\text{and } \varphi = 2\theta$$

where $|K|$ is the absolute value of reflection coefficient, S is the voltage standing wave ratio, and θ is the distance of the voltage minimum from the reference point expressed in electrical degrees.

This chart then shows very clearly the effects of variation of load impedance on both frequency and power output. A feature of great significance which is common to many types of magnetrons is that the region of highest power output is near the region of greatest frequency change. This requires some compromise in choice of working point. The choice may be assisted by putting on the chart other information such as RF bandwidth, and regions of sparking, over-heating, or frequency instability. For the diagram shown, a good working point would lie near the centre, where powers of 60 to 70 watts would be obtained. In this region a change of standing wave ratio from 1·0 to 1·5 would produce a frequency change of $\pm$ 5 megacycles per second at the most.

[13] G. B. Collins (Ed.), *Microwave Magnetrons* (Massachusetts Institute of Technology, Radiation Laboratory Series, Vol. 6), McGraw-Hill, New York, 1948, pp. 40-2, 316-20.

[14] See Chapter VI, section 6.

Rieke diagrams may of course be drawn on any suitable impedance chart. There is no need to stress their importance to the user who requires to minimise frequency variations.

The problem of frequency dependence on impedance is also being attacked from the other direction. Attempts are being made to construct magnetrons which show no such frequency

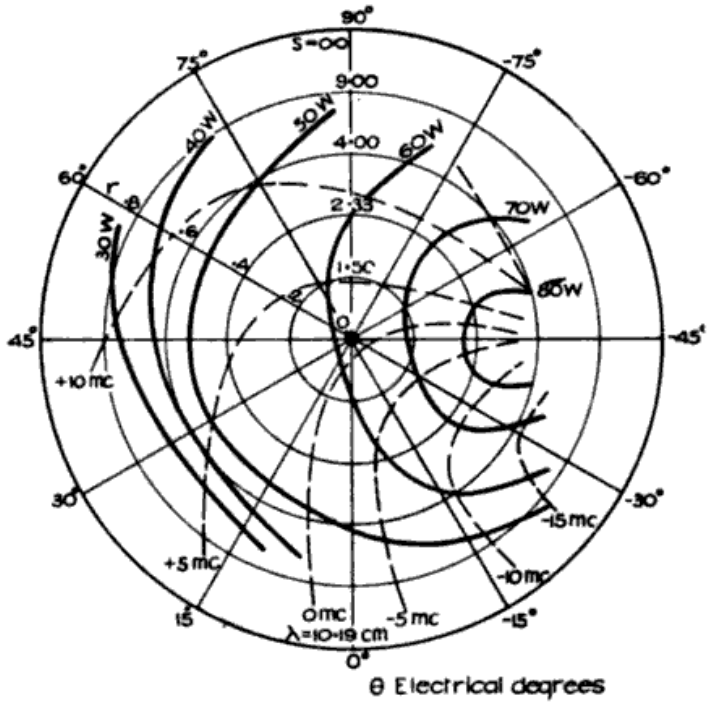


Figure 28.—Rieke diagram showing the effects of varying load impedances, on frequency and power output of a magnetron as measured by standing wave ratios and displacement of the voltage minimum.

criticality when delivering maximum power output, the method of attack being to insert a highly resonant cavity between the load and the tube.

6. General Features of the Practical Resonator Magnetron

In order to get a concrete idea of the practical fruits of development let us examine a successful type, the British CV76, which was made in very large numbers and saw wide service in the war.

The CV76 Magnetron

The operating conditions of this tube under pulse conditions are: wavelength, 10 centimetres; magnetic flux density 0·23 webers per square metre (2300 gauss); peak anode current,

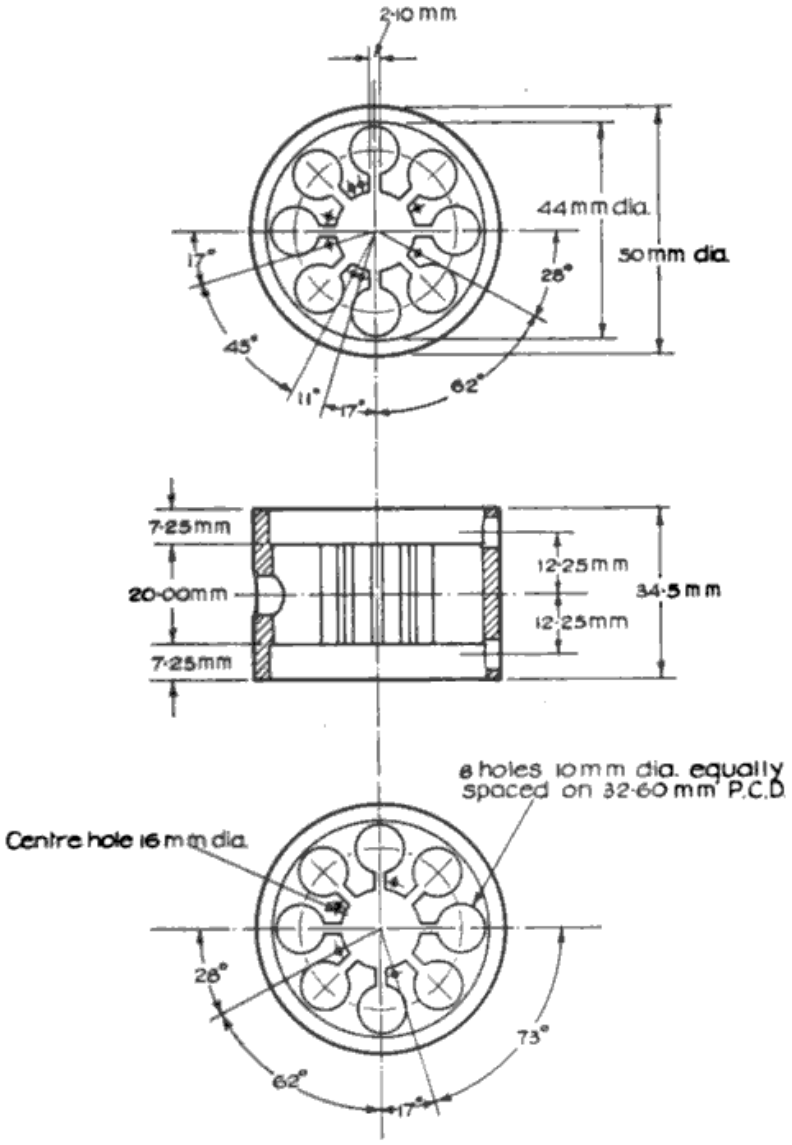
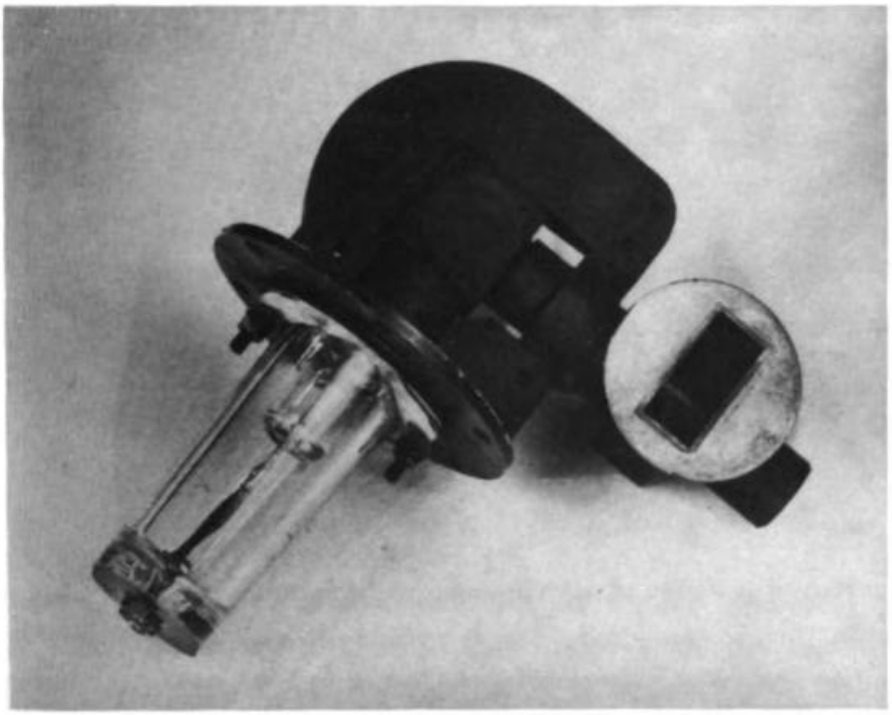
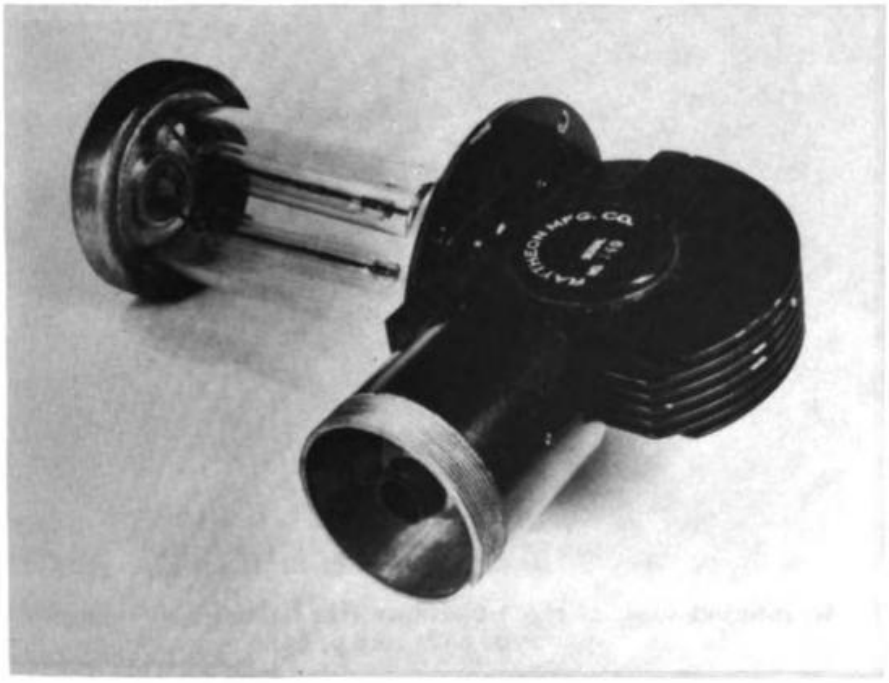


Figure 29.—Anode block used in type CV76 magnetron.

40 amperes; peak power output, 400 kilowatts; efficiency, 50 to 60 per cent; pulse duration, 0·5 microseconds; and



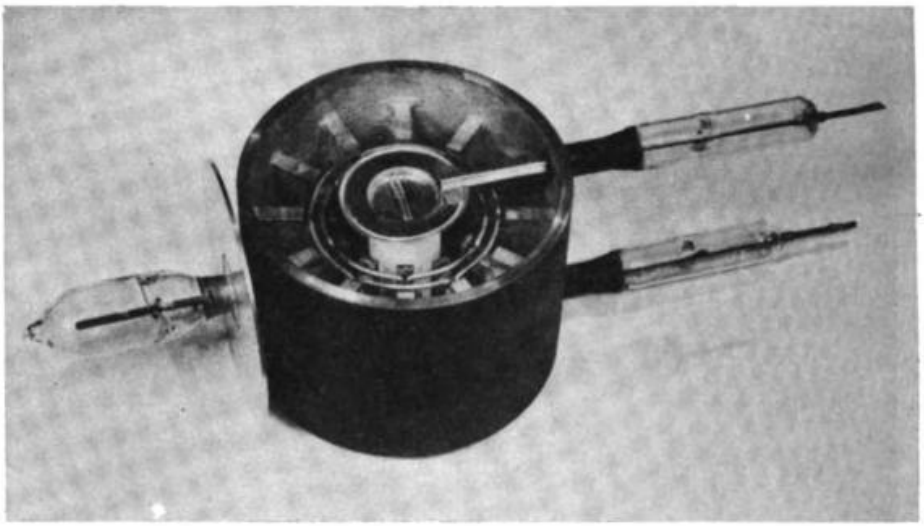
(*a*) The Western Electric Magnetron type 725.



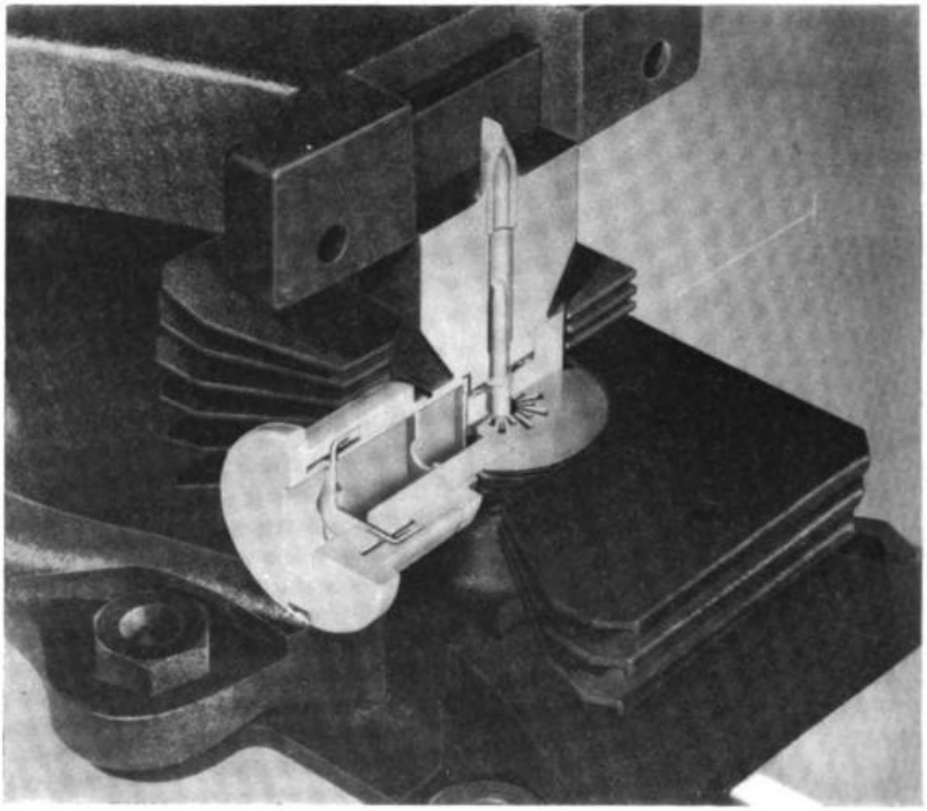
(*b*) The Raytheon Magnetron type HK7 (see p. 64).

PLATE IV.

facing p. 58]



(*a*) Partial assembly of AWV Company magnetron type AV20 (see p. 66).



(*b*) An internal view of the 1·25-centimetre "rising-sun" magnetron type 3J21 (see p. 66).

[*From "Radar Systems and Components", Members of the Technical Staff of Bell Telephone Laboratories, D. van Nostrand Co. Inc.*, 1949.

PLATE V.

pulse repetition frequency 500 cycles per second. The life is of the order of 1,000 hours.

The tube with end-plates removed is shown in Plate III (*b*), along with the anode block. This block contains the resonant circuits and the anode cavity as in the early Birmingham magnetrons, and forms the main valve envelope. A drawing of the block is given in Fig. 29 and in view of the tolerances required, it offers an interesting problem in machining. The first blocks of this kind were turned and jig-bored, but in quantity production the body and the centre anode hole are turned, and the eight concentric holes for the resonant circuits drilled with a jig. The slots are cut in a shaping or slotting machine with proper indexing fixtures, or are broached. The side holes for the output and filament leads are jig drilled and reamed or tapped.

The mode-locking straps are bridges of copper wire, peened into holes drilled in the segments. The strapping used is that referred to above as Y-B strapping and aims at π mode excitation.

The cathode is mounted centrally on the two tungsten filament leads which are sealed to the body by the use of

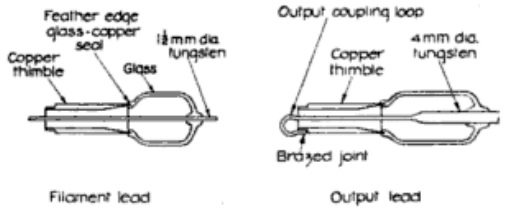


Figure 30.—Arrangements for filament and output leads of a type CV76 magnetron.

feather-edged copper thimbles, glass-copper seals, and glass-tungsten seals. The complete filament lead assembly is shown in Fig. 30, along with the similarly constructed output lead.

The cathode is of nickel of 6 millimetres outside diameter and 0·5 millimetres wall thickness, and is fitted with end-plates. It is coated with the conventional emission mixture of double barium-strontium carbonates which on subsequent exhaust treatment give an oxide coating similar to that widely used on receiving-tube cathodes. This cathode is indirectly headed by an internal spiral of tungsten wire.

In assembly, the filament and output leads are fitted into the anode block and brazed into position. The cathode is then welded onto the filament leads. The end-plates, which are flat copper discs turned to fit the ends of the anode block, are either brazed on, or attached by the extremely useful General Electric Company (Wembley) device of a ring gasket of gold wire. In this latter method, the body with its assembled seals, and plates, and gold rings, is clamped between plates and transferred to the exhaust bench where the gold gasket forms a vacuum tight seal during the baking part of the exhaust process. The brazing processes require the use of a protective gas and chemical washing at various stages in order to keep the interior of the tube clean enough for the exhaust to be effective.

The exhaust process is carried out on a pumping system of the usual kind employing a diffusion pump backed by a mechanical pump. An exhaust schedule includes about an hour's bake at 500 degrees centigrade, followed by cathode activation and the passage of some space current—steady DC or pulsed with magnetic field. A getter is not generally used.

The external fittings are then soft soldered to the exhausted valve. These are the cooling radiator and the output thimble sleeve. In operation the valve is forced-air cooled, an air flow of the order of 50 cubic feet per minute being used. Forced air cooling is in universal use for high power magnetrons—indeed this is one of the advantages of the method of construction initiated at Birmingham.

Magnetron Cathodes

Magnetron cathodes raise several problems of great interest. High power magnetrons demand emission of the order of

10 peak amperes per square centimetre at a wavelength of 10 centimetres, and as much as 30 amperes per square centimetre at 3 centimetres. Those figures are much higher than those normally accepted for oxide cathode design (0·5 to 1·0 ampere per square centimetre). Oxide cathodes are used universally. Their high efficiency made them attractive in the first place and they have subsequently given excellent service. Two factors may contribute to such high ratings:

(*a*) "Back bombardment" is inseparable from magnetron operation, and oxide cathodes at normal working temperatures show marked secondary emission; factors of 5 and greater are possible.[15]

(*b*) Pulse operation allows long recovery periods between short active periods.

It is interesting to note that back bombardment results in dissipation of heat at the cathode which allows most magnetrons to run with their heaters switched off once they are started. On the other hand it presents a problem of adequately cooling the cathode in small high-power magnetrons. In the CV76, therefore, the cathode is made of heavy gauge nickel (0·020 inches) welded at one end to a heavy end-plate which is itself welded to one of the filament leads so that as much conduction of heat as possible takes place.

Experience has shown that a cathode poor in emission can lead to frequency instability and sparking. This has been investigated at the M.I.T. Radiation Laboratory (by Coomes, 1944),[16] where DC pulse emission in the absence of a magnetic field was correlated with magnetron performance. Poor emission, measured in this way, was accompanied by poor performance and provided a useful test for the development of better cathodes.

Much work[17] has been done to produce long life cathodes,

[15] M. A. Pomerantz, "Magnetron cathodes," *Proc. I.R.E.*, Vol. 34, 1946, pp. 903-10; and G. B. Collins, *Microwave Magnetrons*,[13] pp. 517-9.

[16] E. A. Coomes, "The pulsed properties of oxide cathodes," *J. Appl. Phys.*, Vol. 17, 1946, pp. 647-54; and G. B. Collins, *Microwave Magnetrons*,[13] pp. 379, 505-6.

[17] G. B. Collins, *Microwave Magnetrons*,[13] pp. 503-39; and J. B. Fisk, H. D. Hagstrum, P. L. Hartman, "The magnetron as a generator of centimetre waves," *Bell Syst. Tech. J.*, Vol. 25, 1946, pp. 167-348.

particularly in higher frequency magnetrons. It was realised that perhaps the most predominant cause of short life was destruction of small areas on the cathode surface by sparking, which produced intense local heating. The use of nickel mesh on the cathode surface was introduced. The mesh was packed with carbonates which provided a reservoir of oxides, and by improving the conduction of heat away from a hot spot, minimised the effects of the arc. More recent developments along these lines include the use of porous metal sintered layers instead of mesh and the use of a mixture of carbonates and powdered metal (e.g. nickel), with porous layers and with mesh.

The use of end-plates on cathodes is universal. Their purpose is to confine the electrons to the active part of the valve. They can cause trouble if they give rise to electron emission, whether primary or secondary, and at times measures have to be taken to deal with this.[18]

The Output Transformer

A resonator magnetron sets a problem in microwave power transmission, since the power generated in small cavity resonators must be extracted and delivered to a feeder. Starting from a loop on one cavity, the most practicable step is to use a short length of evacuated coaxial line of small diameter. That shown on the CV76 is typical. The small line is connected to a large coaxial or waveguide feeder and as the proper load impedance must be presented to the magnetron an impedance-matching transformer is necessary. A transformer for low-power magnetrons is shown diagrammatically in Fig. 31. The slugs are adjusted to produce maximum power. Another type of output transformer for direct connection to waveguide is constructed so that the inner conductor of the coaxial output protrudes into a waveguide and acts as an aerial.

[18] G. B. Collins (Ed.), *Microwave Magnetrons* (Massachusetts Institute of Technology, Radiation Laboratory Series, Vol. 6), McGraw-Hill, New York, 1948, pp. 379, 537-9.

Magnetrons will be described later which typify more recent developments in which a fixed transformer is built into the tube itself and adjusted in the factory so that direct connection may be made to the feeder.

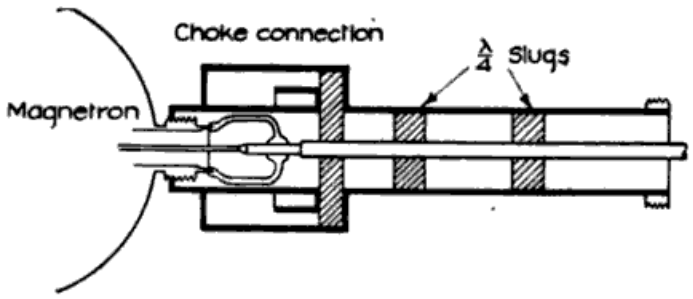


Figure 31.—Coaxial output transformer, for a low-power magnetron.

Much work has been done on magnetron output transformers (for example of this see Collins[19]). Besides presenting a suitable load to the magnetron for which they are designed, they must be mechanically satisfactory and adjustments, if provided, must not be critical.

7. Examples of Magnetron Design

Further information on the design of resonator magnetrons will be found in the literature,[20] together with details of other developments such as tunable magnetrons and so-called "packaged" versions which incorporate a built-in permanent magnet.

In order to convey some idea of the trends in modern design, a description follows of the salient features of a number of magnetrons which have come into widespread use.

A 3-Centimetre Magnetron. Western Electric Type 725

The Western Electric Type 725 is a highly successful 3-centimetre magnetron developed by the Bell Telephone Laboratory.

[19] G. B. Collins, *Microwave Magnetrons*,[13] pp. 481-98.

[20] "Cavity magnetrons," *Electronics*, Vol. 19, 1946, pp. 126-31; J. B. Fisk, H. D. Hagstrum, P. L. Hartman, "The magnetron as a generator of centimetre waves," *Bell Syst. Tech. J.*, Vol. 25, 1946, pp. 167-348; W. E. Willshaw, L. Rushforth, A. G. Stainsby, R. Latham, A. W. Balls, A. H. King, "The high-power pulsed magnetron. Development and design for radar applications," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 985-1005.

Its general appearance is shown in Plate IV (*a*) and its performance is as follows : wavelength, 3·2 centimetres ; magnetic flux density, 0·55 webers per square metre (5500 gauss) ; peak anode current, 10 amperes ; peak anode potential, 12 kilo-

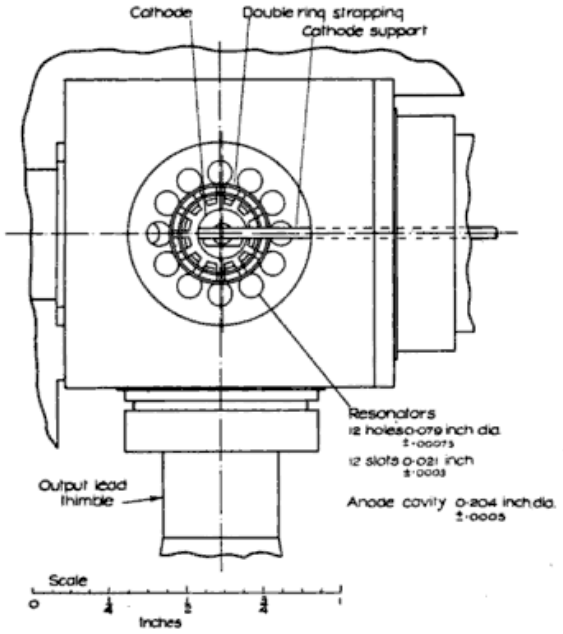


Figure 32.—The internal construction of the Western Electric Magnetron type 725.

volts ; peak power output, 40 kilowatts ; efficiency, 30 to 40 per cent ; duty cycle, 0·001.

A drawing of the anode block is shown in Fig. 32, and attention is drawn to the dimensions of the resonators and anode hole. It is obvious that a formidable manufacturing problem is presented by this minute block, yet this valve was

put into mass production during the war. Double ring strapping is used at each end to ensure π-mode excitation. The cathode, which is indirectly heated by an internal heater, is turned from nickel rod and has a covering layer of nickel mesh impregnated with the usual oxides.

This magnetron has a built-in output transformer which is adjusted and fixed at the factory so that direct coupling can be made to rectangular waveguide with the assurance both of good power output being passed into a reasonably terminated guide, and good frequency stability.

The output transformer is shown diagrammatically in Fig. 33 from which it will be observed that the output lead, which

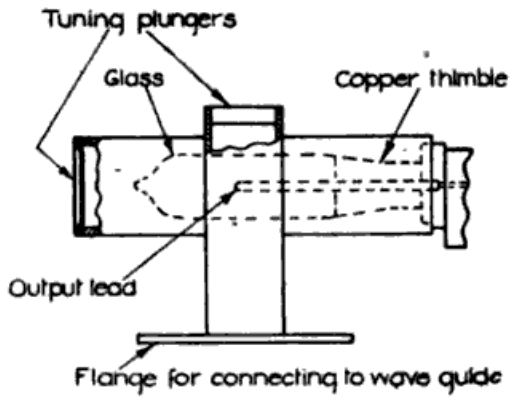


Figure 33.—Output transformer of type 725 magnetron.

acts as an aerial in the waveguide, is vacuum enclosed. This is a simple arrangement which avoids corona losses at the lead and allows one vacuum seal to be dispensed with.

Finally it is noteworthy that very thorough mechanical protection is afforded the fragile copper-glass seals by both output transformer and glass housing round the filament leads. This is a feature of great practical importance.

A High Power 10-Centimetre Magnetron : Type HK7

The HK7 is a high power magnetron which has been made in quantity by the Raytheon Manufacturing Company, and its construction is shown in Plate IV (*b*) and Fig. 34. Approximate

performance figures are : wavelength, 10 centimetres ; magnetic flux density, 0·22 webers per square metre (2200 gauss) ; peak anode current, 40 amperes ; peak anode potential,

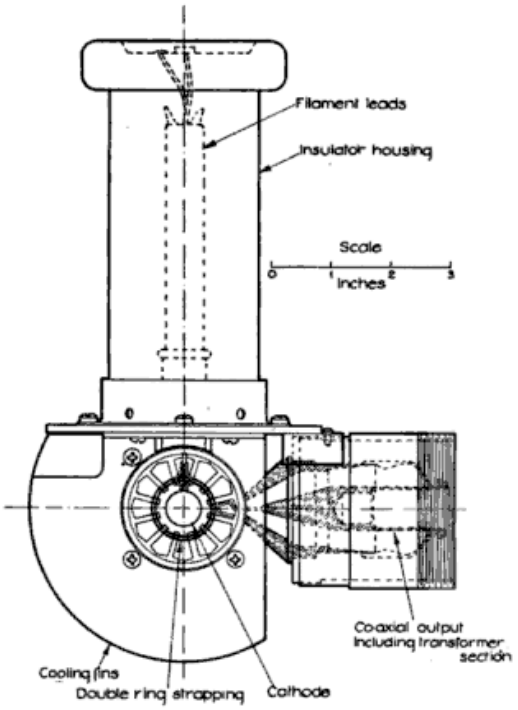


Figure 34.—Construction of the Raytheon Magnetron type HK7.

30 kilovolts ; peak power output, 750 kilowatts ; efficiency, 60 to 65 per cent.

Recent work on a valve with the same basic design as this one has produced power outputs of no less than 3·5 megawatts peak.

Points of interest in the design are: *The coaxial output* which is built to carry high power and incorporates a transformer section which gives a proper match to a 50-ohm line. Mechanically, this output connection is excellent and the glass work is well protected. *The resonators* whose peculiar shape is associated with the interesting method of constructing the anode block from stamped-out sheets approximately 0·060 inches thick, a number of which is piled together and brazed to form a solid whole. This method avoids fine machining of each block and is therefore well suited to mass production. *The filament leads* which are very long to prevent flash-over. They are also provided with corona rings and are well protected by a robust insulator housing.

A 25-Centimetre Magnetron: Type AV20

Plate V (*a*) shows a partial assembly of an Australian magnetron which was designed in the Physics Department, University of Melbourne, and produced by the Amalgamated Wireless Valve Company. It is of the vane type, the resonator assembly being made up of copper vanes brazed to a copper shell. Double ring strapping is used, the strapping rings being formed by copper wire bridges. The design as a whole is orthodox and permits the valve to be made entirely from local materials. In its final form it is fitted with cooling fins and a filament lead housing.

A 1·25-Centimetre "Rising-Sun" Magnetron: Type 3J21

Plate V (*b*) shows some details of the 3J21 magnetron which uses the "rising-sun" anode block. This system of anode cavities gives good mode separation without straps, and provides a practical construction for the minute blocks necessary for magnetrons operating at very short wavelengths. The cavities are made out of a solid block of copper by a single hobbing operation. The waveguide output coupled to the resonators by a "window" transformer should also be noted.

CHAPTER IV

TRIODE POWER OSCILLATORS

ALTHOUGH easily the most outstanding valve developed for radar purposes has been the resonant magnetron, triodes have nevertheless played an important part as oscillators in pulse transmitters. Their contribution has been on the longer wavelengths associated with radar and in this role they are unlikely to be replaced. Since the development of triodes for very high frequency operation was not a radar development and since they are already described fully in the literature, this chapter concerns itself only with those aspects of valve design which are peculiar to pulsed operation. These are discussed first in relation to the electrical design of the valve itself and then from the point of view of the reaction of this design on the geometry of the valve and its associated circuits.

1. Electrical Design

It is fortunate that pulsed operation not only requires but permits high peak powers to be obtained from the triode valves, often at frequencies higher than is possible in continuous wave operation.

Three interrelated factors enter into production of high powers ; the cooling of the valve, the applied DC voltages and the emission of the cathode. A direct result of operating the valve over only a short portion of the total time is that the mean power to be dissipated is reduced much below the peak value, the factor being generally of the order of 1,000. Hence a given valve can handle peak powers many times greater than its normal rating, provided always that it can be made to generate these higher peaks of power. In order to increase

the power output, the anode voltage and/or current must be increased. Even with conventional designs, it is possible to increase the voltage as much as 4 times and the current by a factor of perhaps 2, giving a power increase of 5 or 6 times.

This factor can be increased even more by modification of the valve design. A good vacuum and the elimination of all unnecessary insulating material within the valve ensure that high voltages will cause little internal trouble. High voltages can be tolerated external to the valve because the ions associated with incipient corona have time to disperse or can be forcibly dispersed by air blast during the relatively long rest periods between pulses. Full use can be made of this by pulsing the power supply to the valve anode—by so-called anode modulation.

It was originally thought that tungsten filaments were necessary for operation at high anode voltages, and attempts to increase the emission current were then limited by grid heating. Fortunately it was found that oxide coated cathodes could withstand quite high anode voltages under pulsed conditions and a tremendous gain in emission followed immediately. An additional advantage was the reduction in filament power which decreased the total power loss to be dissipated ; because it is continuous, the filament load always represents a considerable portion of the total.

Even with the great increase in peak power output, no very special demands are made by the cooling requirements which can always be met by the provision of radiator fins and an air blast. The overall result is that in spite of the high peak power available the valve is quite small in size, a further advantage for very high frequency operation.

2. Physical Design

The main effort put into the design of valves for radar purposes has been in the direction of increasing the upper limit of frequency and of increasing their efficiency near this limit. The main factor limiting the high frequency operation of triodes is electron transit time which in turn is related to the

size of the valve and the applied potentials. It has been shown above that pulsed operation permits a decrease in size and an increase in applied voltage both of which raise the frequency limit.

The need for improved efficiency has resulted in designs which have developed along the following lines. At the desired frequencies, the interelectrode capacities within the valve represent impedances which are very much smaller than those considered optimum in low frequency practice. In the first place they can give rise to high currents and therefore high losses, and then they require small tuning inductances which are not easily realised. The capacities should therefore be kept small and of such a construction that they can be connected to the remainder of the circuit by leads of low inductance. For the same reasons no more circuit elements than are absolutely necessary for proper operation as an oscillator should be added.

A design procedure[1] is available for determining the circuit elements which should be added to give the desired frequency, output and control conditions. It remains to find a suitable physical form for these elements. The principal element is always inductive and the inductor which has the greatest physical size for a given value of inductance is of the transmission line type, which may take the form of a twin or coaxial line or at the highest frequencies may be a radial line or cavity. The valve structure must permit a simple connection to such elements and it has therefore taken the concentric form which is well illustrated by the NT99 valve (Plate VI (*a*)). This valve is capable of a peak power output of 40 kilowatts at frequencies in the vicinity of 600 megacycles per second; the anode voltage is 8,000 and the filament heating power is 40 watts. It is typical of many produced by the Research Laboratory of the General Electric Company, Wembley, which played an outstanding part in the development of such valves for radar purposes.

[1] O. O. Pulley, "Ultra high frequency oscillators," C.S.I.R. Radiophysics Laboratory, Report RP 92, 1941.

An alternative construction popular at lower power levels is the plane-element type with disc seals of low inductance. It is exemplified by the CV90 and the GL446 described in Chapter XI.

In both these types the inductance of the lead from the electrode to a point external to the valve where the circuit can be connected is kept as low as possible. This allows a high proportion of the inductive tuning element to be external to the valve and hence under control. Known artifices for increasing the size of the external circuit such as the use of a line length greater than half a wavelength are permissible but seldom satisfactory.

The relatively simple method of increasing power output by the use of several valves in parallel or push-pull is much used and has a bearing on valve design because of the length and hence impedance involved in the inter-connections. The design procedure quoted above shows that in push-pull opera-

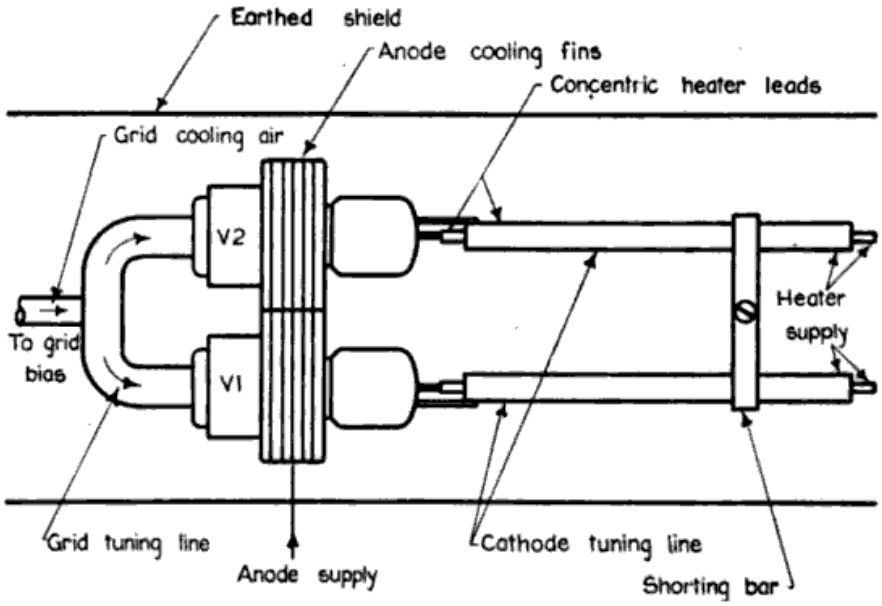


Figure 35.—Typical schematic arrangement of a push-pull oscillator for 500 megacycles per second.

tion the impedance between one pair of corresponding electrodes should preferably be zero or very small, and this can be achieved in a valve such as the NT99 by strapping the two

anodes directly together. The grid and cathode circuits then become part of a twin transmission line system. To prevent radiation they should be shielded by an extension of the anode system.

Fig. 35 shows schematically the arrangement of a push-pull oscillator suitable for operation at about 500 megacycles per second while Plate VI (*b*) illustrates the construction of a transmitter for a somewhat lower frequency. In the latter the grid and cathode tuning lines run backwards from the valve assembly while the valves themselves are mounted on a closed box of insulating material through which the cooling air is blown.

Coupling to the load can be effected by tapping off either the grid or cathode lines, the latter usually being more convenient. Loop coupling can also be used and in some cases is necessary in order to isolate the oscillator from its load.

Many variations in arrangement, construction, etc. have been described in published literature or can readily be devised to suit particular conditions.

CHAPTER V

MODULATORS

RADAR transmitters are required to radiate pulses of radio frequency for periods from 0·1 to 20 microseconds repeated at rates up to 2,000 pulses per second. Triode oscillators are used for transmitter frequencies up to about 1,000 megacycles per second, while for higher frequencies magnetrons are generally employed. The usual method of modulation is to apply a high voltage of rectangular waveform to the anode of the oscillator if it is a triode, or to the cathode if a magnetron. Provided the build-up and decay times are short compared with the duration of the pulse, the resultant current waveform and radio frequency envelope are similar to the applied pulse. This method of modulation corresponds to anode modulation of radio telephone practice and, as explained in Chapter IV, it allows tubes to be run at higher anode voltages and consequently with greater power output than is normal. Grid modulation of triode oscillators is seldom used because in this method the anode potential has to be applied continuously and full advantage cannot be taken of the virtues of pulse operation.

Because power is transmitted only during a small fraction of the total operating time, peak powers which are extraordinarily high compared with those of radio communication practice are achieved with moderately sized radar equipment. That fraction of the total time during which the transmitter actually operates is called the " duty cycle," and is equal to the product of pulse duration and repetition frequency. Peak values of quantities such as power can then be converted to average values by multiplying by the duty cycle. For example, a high power air warning set employing a 4 microsecond pulse at a repetition frequency of 250 pulses per second may use a magnetron accepting a peak input power of one megawatt. The duty cycle is 1/1000, so that the mean power input to the magnetron is only one kilowatt.

In practice, the output waveform of a modulator may not have the ideal rectangular shape and the permissible departure

depends on the characteristics of the oscillator. A variable anode voltage will cause both magnetron and triode oscillators to change frequency, so that the receiver will not be in tune during some portion of the pulse. In addition magnetrons driven by a varying voltage may jump discontinuously to a different frequency, a phenomenon known as " double moding." This may also be caused by a rapid rise of the leading edge of the pulse and, to prevent it, modulators delivering flat-topped trapezoidal pulses have been proposed.

In the oscillating region, magnetrons have approximately constant voltage characteristics so that any modulator will produce a more rectangular voltage pulse with a magnetron than with a resistive load. The reverse is true for the current pulse.

The techniques described in this chapter have application in fields other than radar, consequently the needs of those who have not previously had access to this information have been kept in mind.

1. Basic Modulator Types

Radar modulators may be divided into two main types. The first applies a high tension supply to the oscillator anode for the duration of the pulse and then disconnects it, as illustrated in Fig. 36. The " on-off " switch is a high vacuum triode or

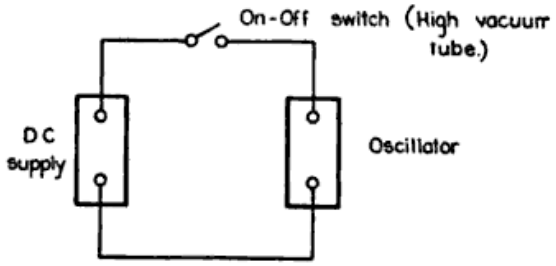


Figure 36.—Basic high vacuum tube modulator circuit.

tetrode which operates under particularly arduous conditions. Alternatively the modulator may be thought of as a power amplifier developing the high voltage pulse across a load,

which is the oscillator in Fig. 37. Because of the difficulty of manufacturing suitable switch tubes, and because they operate at high voltage and low current, this type of modulator has been largely superseded. It is referred to as a "high vacuum tube" or "hard tube" modulator.

The second type of modulator uses a pulse forming network containing capacitors, the total energy of which is discharged into the oscillator by a switch which remains closed until after the discharge is complete. The switch must pass from the non-conducting to the fully conducting state in a time which is short compared with the pulse duration, but the only requirement as to its rate of opening is that it should become non-conducting in a time short compared with the interval between pulses. The ability of an open spark in air to perform this operation has led to the adoption of this type of modulator for many applications.

The basic pulse forming network is the loss-free open circuited transmission line, which is charged and connected at one end to a resistance equal to its characteristic impedance. It produces across the resistance a rectangular voltage pulse of magnitude equal to half the voltage to which the line was initially charged. The pulse terminates when the disturbance caused by the closing of the switch arrives back at the load after reflection at the open end. The waveform is determined by the network and not by the switch. The interval between pulses is available for recharging the network so the demand on the power supply is not severe. This type of modulator is referred to as a "line type modulator."

The successful development of voltage step-up transformers capable of handling high power pulses constituted a notable advance in modulator technique. Magnetrons usually operate at inconveniently high voltages of 15 to 40 kilovolts, and the provision of insulation to withstand this or a greater voltage in the power supplies and elsewhere adds considerably to the size of the equipment. With the introduction of "pulse transformers," as they are called, it became practicable to operate the main part of the modulator at the most convenient voltage and restrict the high voltage to the output terminal

A further feature of the use of pulse transformers is that they permit the separation of the modulator[1] from the transmitter, pulses being transferred by means of relatively low voltage coaxial cable. The insertion of a cable does not cause distortion of the pulse if the terminating impedances are equal to the characteristic impedance of the cable. Magnetron input impedances are about 500 to 1,000 ohms while the characteristic impedance of convenient coaxial cable is about 50 ohms so that by connecting the magnetron through a transformer of about 10/1 impedance ratio it is possible to match it to the cable. In the case of a line type modulator the pulse forming network is made equal in impedance to the coaxial cable and only one transformer is employed. A high vacuum tube modulator requires a transformer at each end of the cable since the modulator is itself a high impedance device.

In comparing the two basic modulator types the line type modulator has two salient advantages. It has a simpler switch and when used with a pulse transformer a much lower supply voltage. The high vacuum tube modulator requires a supply voltage some 15 per cent greater than the output pulse voltage, while a typical line type modulator employing DC resonant charging and a 4 : 1 step-up pulse transformer requires a supply voltage only one quarter of the output. The line type modulator can also be made simpler, smaller and more efficient. On the other hand the hard tube modulator[1] is more flexible and is therefore often used for laboratory purposes.

2. High Vacuum Tube Modulators

The basic high vacuum tube modulator circuit shown in Fig. 36 cannot be used in practice because the cathode of the switching tube is at high DC potential. Two practical modifications are shown in Fig. 37 (*a*) and (*b*), the first being transformer coupled and the second capacitor coupled. The grid

[1] R. H. Johnson, "Hard-valve pulse modulators for experimental use in the laboratory," *J. Instn. Elect. Engrs.*, Vol. 93, Part III A, 1946, pp. 1043-57.

of the switch tube is biased beyond cut-off during the quiescent period and driven positive rendering the tube highly conducting during the pulse.

If the pulse applied to the grid is exactly rectangular, distortion still occurs in the output in the manner illustrated

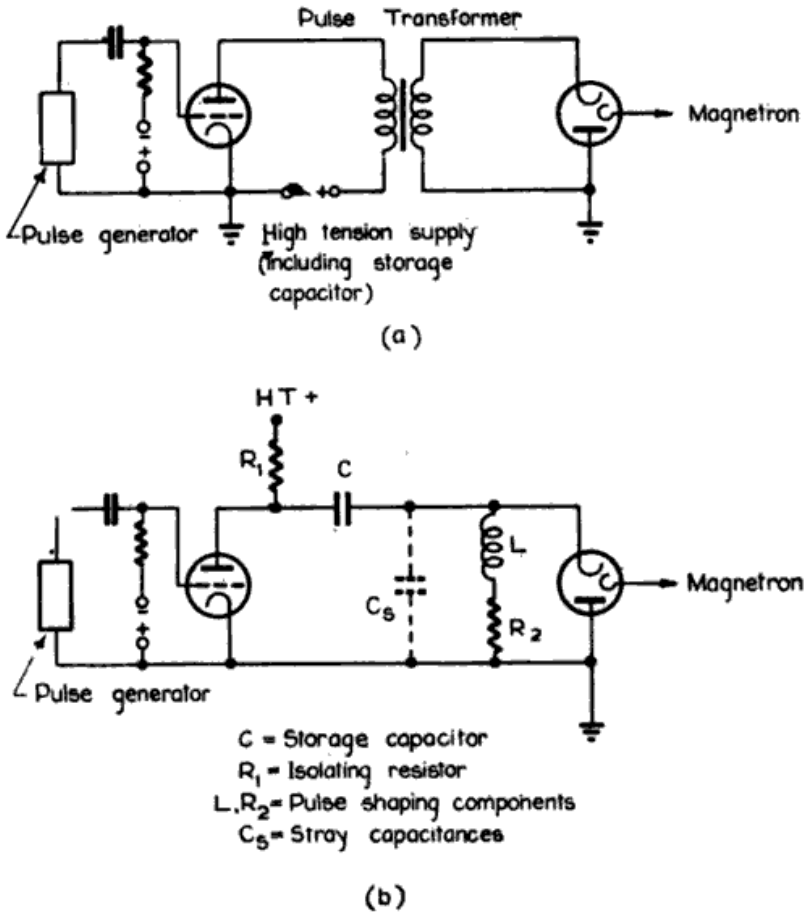


Figure 37.—High vacuum tube modulator circuits. (*a*) Transformer coupled. (*b*) Capacitor coupled.

in Fig. 38. The initial rise of the pulse is delayed while the stray capacitances C_s in Fig. 37 (*b*) are charged through the switching tube. The voltage falls during the pulse because the coupling capacitor C is being discharged by the current through the magnetron and the components L and R_2. At the end of the pulse the switching valve becomes non-conducting and after the voltage has fallen about 10 per cent the magnetron also be-

comes non-conducting. The stray capacitances then have to discharge through L and R_2. If the values of these components are properly chosen the discharge can be critically damped. If L is replaced by a high resistance a slow decay results. If L only is used an oscillation with C_s as capacitance is set up. Rapid damping of the oscillation may be obtained by connecting a low impedance diode in parallel with the magnetron with such a polarity that it conducts when the magnetron anode is negative with respect to its cathode. This produces heavy damping while the diode is conducting but wastes no power when the pulse is of opposite polarity. The arrangement suffers from the disadvantage that it adds another major component to the system. Fig. 38 shows the after pulse effects of the various components mentioned.

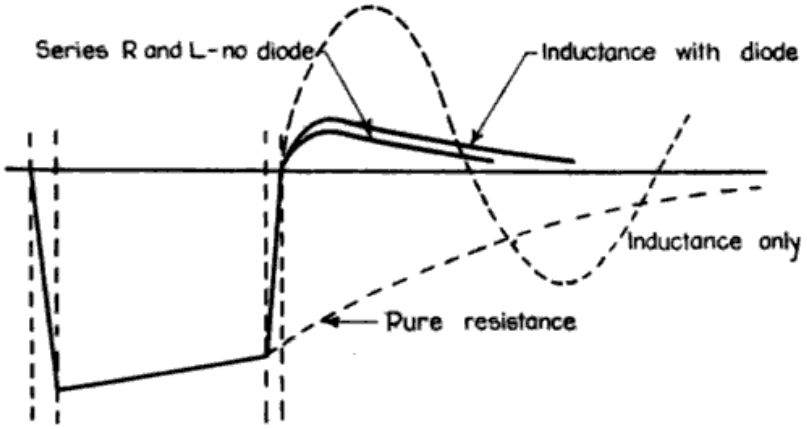


Figure 38.—Magnetron voltage pulse with various damping methods.

The positive pulse required to drive the grid of the switch tube is large and difficulty is experienced in obtaining it from a driving tube of reasonable dimensions. The difficulty arises because, in order to economise in size and power consumption, high power tubes are arranged to conduct only during the pulse. Such tubes give negative pulses to a load in the anode, and to

obtain positive pulses it is necessary to couple to the cathode. With cathode coupling a large grid driving voltage, greater than the output voltage, is required. In the widely used " bootstrap " modulator circuit (Fig. 39) this difficulty is overcome by

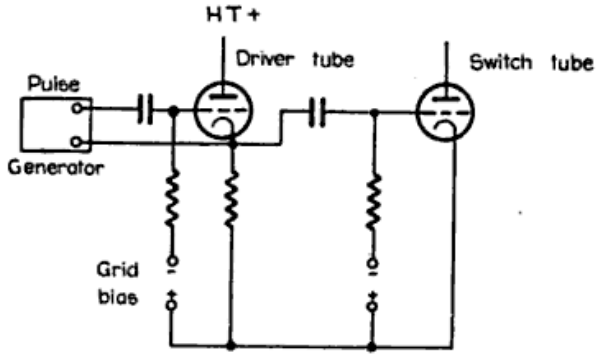


Figure 39.—Bootstrap driver circuit.

connecting a pulse generator between cathode and grid of the driver tube and arranging for the whole of the pulse generator circuits to fluctuate in potential with the driver cathode.

High Vacuum Modulator Tubes

Since the pulse required to drive a typical magnetron is of the order of 30 amperes at 20 kilovolts the switch is required to withstand a similar voltage when non-conducting and to pass a similar current when conducting. The current must be attained with a low voltage drop across the switch tube and it is desirable that the requisite grid swing between the conducting and non-conducting states should be low.

These characteristics are difficult to attain simultaneously in a tube of small size and heater power, and the successful development of such tubes has demanded extremely good high vacuum technique. Table 3 lists the characteristics of some of the tubes used for this purpose.

TABLE 3—CHARACTERISTICS OF SOME TYPICAL HIGH VACUUM MODULATOR TUBES

Tube Type		*Cathode*	*Cathode power watts*	*Maximum anode voltage*	*Peak anode current amperes*
NT100	tetrode	oxide	48	12,000	10
829A	tetrode	oxide	14	5,000	10
715B	tetrode	oxide	58	15,000	15
304TH	triode	thoriated	130	18,000	7.5
6C21	triode	thoriated	150	33,000	15

3. Line Type Modulators[2]

The essentials of the line type modulator are the pulse forming network, the circuit for charging it, and the switch which discharges the network into the load. Rotary spark gaps, fixed triggered gaps both open and closed, and thyratrons have found application as switches, and the comparative ease with which satisfactory switches can be obtained has, among other reasons, resulted in the line type modulator almost completely superseding the high vacuum tube modulator.

Theory of Discharge of a Transmission Line

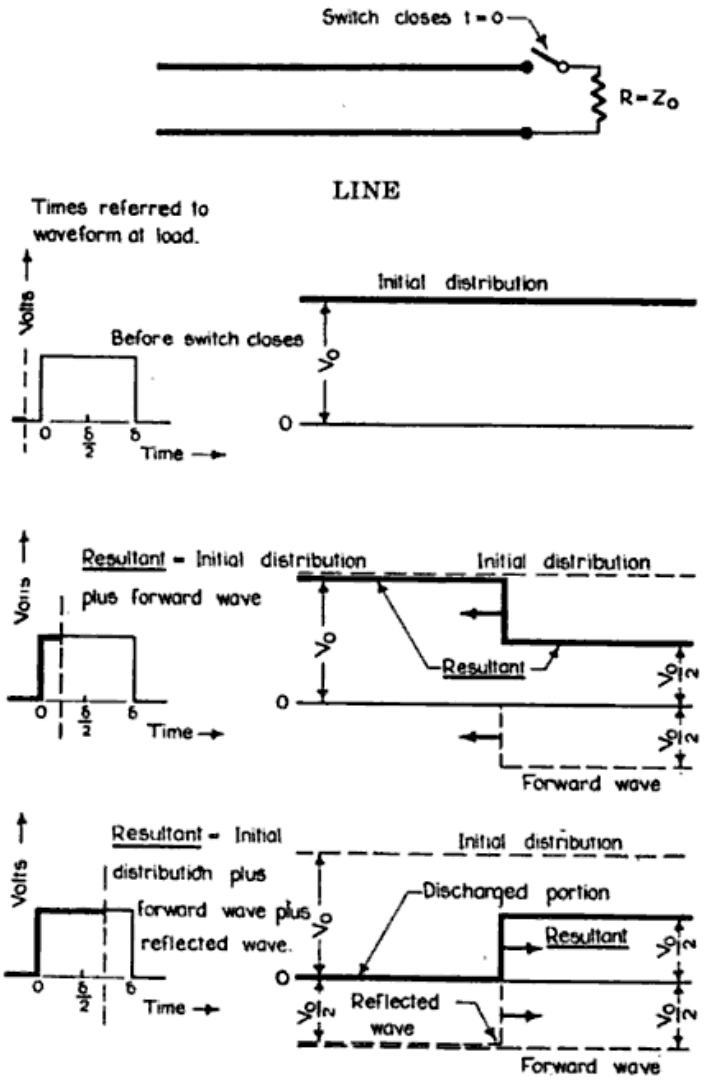
The development of the pulse forming network followed on the realisation that a uniform, low-loss, open-circuited transmission line gives a rectangular pulse when discharged into a resistance equal to its characteristic impedance.

Consider the case of Fig. 40 in which a line of characteristic impedance Z_o is charged to a voltage V_o and connected to a resistor R of magnitude Z_o at time $t = 0$. For a short interval the line behaves as a generator of electromotive force V_o and internal impedance Z_o applying a voltage $V_o/2$ to the load. The initial voltage drop from V_o to $V_o/2$ gives rise to a rectangular wave of voltage $V_o/2$ travelling away from the load. This travelling wave is completely reflected without change of sign at the open end and travels back to the load, where it

[2] K. J. R. Wilkinson, "Some developments in high-power modulators for radar," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 1090-112.

reduces the voltage to zero. The system is then completely discharged, the duration of a pulse being equal to twice the time of travel of a wave along the line.

If R is not equal to Z_o, the discharge will reach completion



TRAVELLING WAVES

Figure 40.—Discharge of a matched transmission line showing travelling waves.

asymptotically. To find the conditions at any given time the steady voltage originally present and all travelling waves must be superposed.

Let V be the voltage across R. Immediately after closing the switch only one travelling wave of magnitude $V_o - V$ is present. The line current $(V_o - V)/Z_o$ equals the load current V/R so that

$$V = \frac{R}{R + Z_o} V_o.$$

This state persists until time $t = \delta$ when the wave reflected from the open end reaches the load. This wave of magnitude $V_o - V$ or $\frac{Z_o}{R + Z_o} V_o$ is partially reflected at the load, the reflection coefficient being $\frac{R - Z_o}{R + Z_o}$, so that in the interval $\delta < t < 2\delta$

$$V = \left(\frac{R}{R + Z_o} - \frac{Z_o}{R + Z_o} - \frac{Z_o}{R + Z_o} \cdot \frac{R - Z_o}{R + Z_o} \right) V_o$$

$$= \frac{R(R - Z_o)}{(R + Z_o)^2} V_o.$$

Continuing the calculation, a series of decreasing steps is

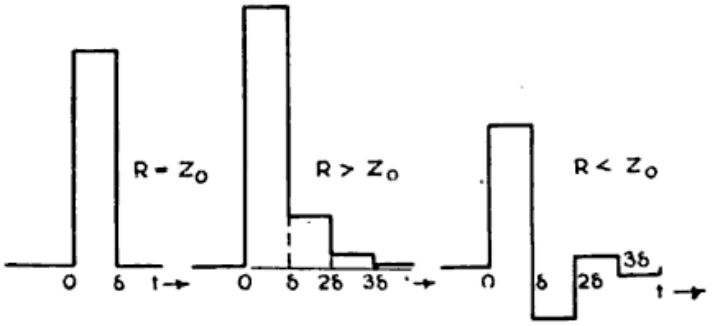


Figure 41.—Discharge of transmission line into resistance loads of different magnitudes.

found, which if $R > Z_o$ are of one sign, or if $R < Z_o$ are of alternate signs as in Fig. 41.

If the load is not a fixed resistance, but has a non-linear voltage-current characteristic independent of the rate at which

it is traced out, a stepped waveform will still result. This case, which may be treated graphically, is important in view of the marked non-linearity of magnetron characteristics.

Alternatively, the closing of the switch may be regarded as the application of a step-shaped voltage between its contacts. The alternating current impedance of the network which gives a current response of rectangular waveform when such a voltage is applied may be calculated by a well-known theorem.[3] It transpires that a uniform open circuited transmission line in series with a load resistance equal to the characteristic impedance is required.

Networks Equivalent to the Uniform Line[4]

An actual transmission line to produce pulses of the length commonly used in radar (1 microsecond or more) would be of inconvenient dimensions. Hence networks consisting of lumped elements are used to approximate the performance of the open circuited line.

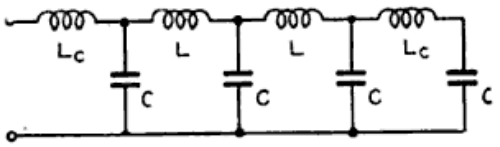


Figure 42.—Low pass filter pulse forming network.

Low Pass Filter Pulse Forming Network. The simplest and most commonly used equivalent network is the low pass filter shown in Fig. 42. Constructional advantages of this network

[3] The current response $A(t)$ of a network to a unit step-shaped applied voltage is related to the corresponding alternating current impedance $Z(j\omega)$ by the equation

$$\frac{1}{aZ(a)} = \int_0^\infty e^{-a\lambda} A(\lambda) d\lambda$$

in which a is real and positive. Assuming a unit current pulse of duration δ for $A(t)$, we have $Z(j\omega) = \frac{1}{2} + \frac{1}{2} \coth j\omega\delta/2$.

[4] E. L. C. White, "The use of delay networks in pulse formation," *J. Instn. Elect. Engrs.*, Vol. 93, Part III A, 1946, pp. 312-4.

are that all capacitors have the same value and the inductors can be wound as a continuous coil on a single former and tapped where required. Two or more of these networks may be connected in tandem to give a choice of pulse duration.

The following relations are satisfactory for design purposes

$$Z_o = \sqrt{\frac{L}{C}}$$

$$\delta = 2n\sqrt{LC}$$

where Z_o = characteristic impedance of network at zero frequency (ohms), δ = pulse duration (seconds), L = inductance per section (henries), L_e = end inductance (henries), C = capacity per section (farads), n = number of sections. Empirically it is found that good pulse shapes can be obtained with 3 or 4 sections when $L_e/L = 1{\cdot}1$ to $1{\cdot}2$.

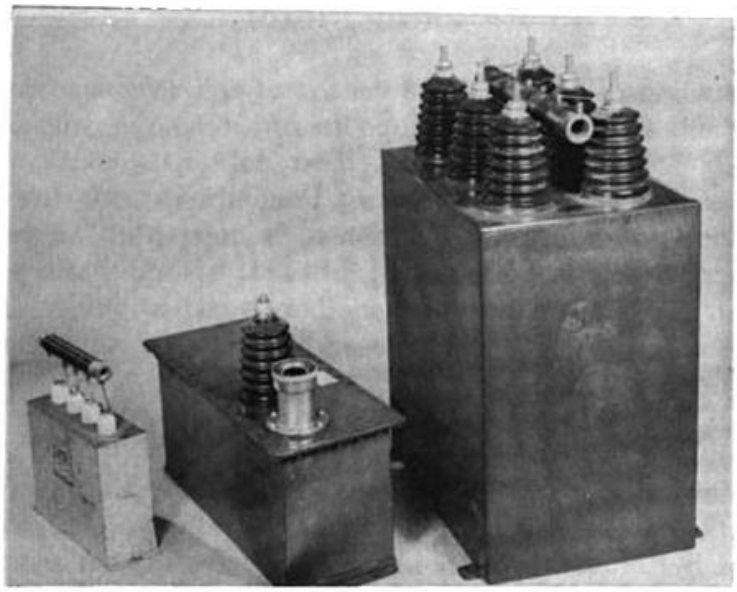
For high voltage networks the capacitor and coil assembly is usually placed in a sealed tank with terminals brought out as required. Oil impregnated paper dieletric is most common, although for some low voltage networks special organic-inorganic dielectric capacitors moulded into a solid block have been developed. Plate VII (*a*) shows some typical low pass filter networks.

Bartlett Pulse Forming Network. Bartlett of the General Electric Company obtained several possible pulse forming networks by expanding the expression for the input impedance and admittance of a uniform open circuited transmission line in various ways. Only the circuit which has found most application will be described here.

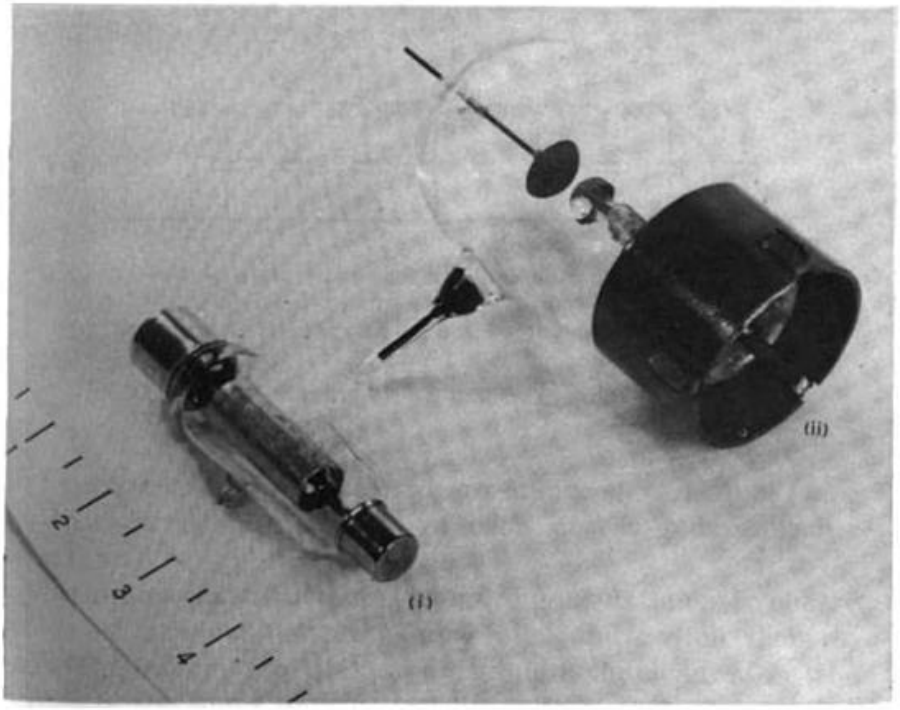
The input impedance of the open circuited transmission line of characteristic impedance Z_o in series with its characteristic impedance is

$$Z_{IN} = Z_o + Z_o \coth \frac{j\omega\delta}{2}$$

where ω is the angular frequency and δ is the time taken for a



(*a*) Pulse forming networks.



(*b*) Enclosed fixed triggered gaps. (i) 1B22. (ii) CV85 (see pp. 89, 90).

PLATE VII.

facing p. 84]

wave to travel to the far end and back. In terms of the partial-fractions expansion of the hyperbolic cotangent, viz.

$$\coth x = \frac{1}{x} + \sum_{n=1}^{\infty} \frac{2x}{n^2\pi^2 + x^2},$$

we have

$$Z_{IN} = Z_o + \frac{2Z_o}{j\omega\delta} + \sum_{n=1}^{\infty} \frac{j\omega\delta Z_o}{n^2\pi^2 - \omega^2\delta^2/4}.$$

The first term is the series resistance Z_o, the second is the reactance of a capacitance equal to $\delta/2Z_o$, and the nth term of the series represents the reactance of a parallel combination of a capacitance $\delta/4Z_o$ with an inductance $Z_o\delta/n^2\pi^2$. Fig. 43 illustrates the equivalent circuit.

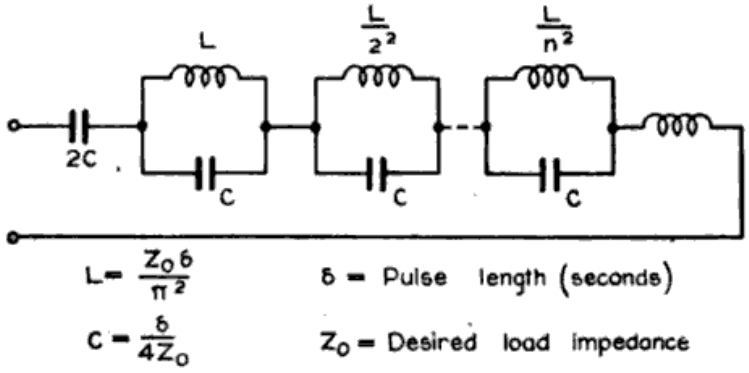


Figure 43.—Bartlett pulse forming network.

The large storage capacitor $2C$ must withstand twice, and the others $2/n\pi$ times the pulse voltage under normal conditions. If for any reason the transmitter becomes short circuited, however, the voltage across the small capacitors will reach twice the quoted values and spark gaps may be needed for protection. In practice four or five parallel circuits are used and a series inductance added, its value being determined experimentally. This network has the merits that

only one capacitor has to withstand the full voltage and the pulse forming capacitors are all equal in value. It is therefore suitable for construction from standard components and is convenient for experimental purposes. However, these features are not important in a design intended for production, and the disadvantage that the pulse forming capacitors contribute nothing to the stored energy renders the Bartlett line inferior to the low-pass filter network.

Special Arrangements for Saving Insulation

The arrangements described above result in a pulse voltage equal to half the voltage to which the network is charged, so that the network must be insulated to withstand twice the voltage of the output pulse. The Blumlein circuit shown in Fig. 44 gives a pulse voltage equal to the voltage to which

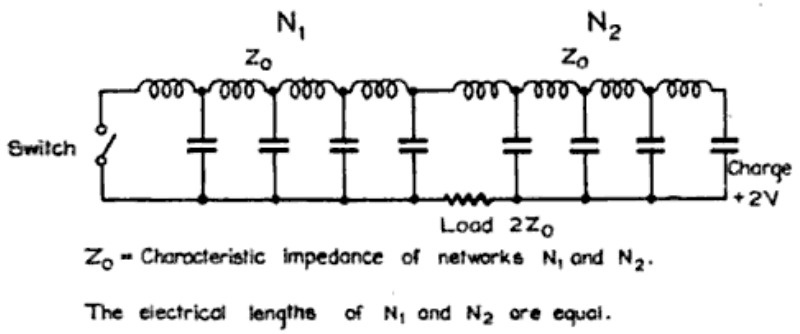


Figure 44.—Blumlein Circuit.

the lines are charged. The load impedance should equal twice the characteristic impedance of each of the pulse forming networks.

Another useful method of economising insulation is to charge two pulse forming networks in parallel and switch them to series connection for discharge through the load after the fashion of the Marx high voltage impulse circuit, as indicated in Fig. 45. Switch No. 1 operates first and the sudden rise

in voltage fires switch No. 2. The "hold-off" chokes must have sufficient inductance to prevent appreciable current building up in them during the pulse. Arrangements such as these have been superseded by the use of pulse transformers but they may still be found useful for experimental purposes when a transformer is not available.

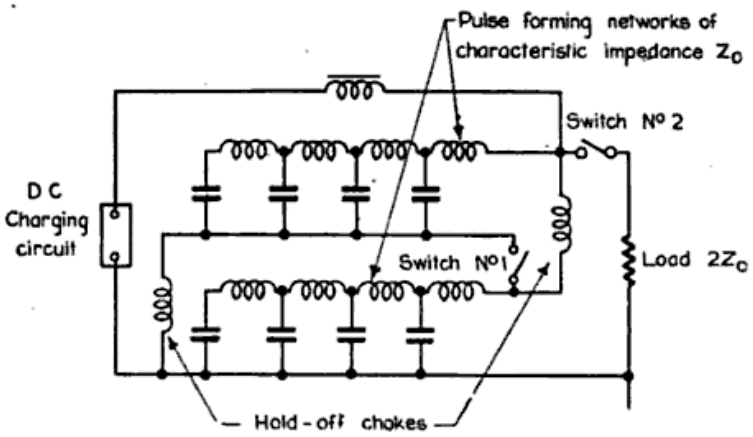


Figure 45.—Parallel charge, series discharge, circuit.

Switches Used in Line Type Modulators[5]

As mentioned earlier the requirements for the switch in a line type modulator are that it should close quickly and remain closed for the duration of the pulse. It should then open in time to permit the charging of the network for the next pulse.

Rotary Spark Gap. The rotary spark gap consists of a set of rotating electrodes which pass close by one or more fixed electrodes. Near the point of closest approach a spark passes which closes the switch. Current ceases on the discharge of the network and the electrodes separate, opening the switch before the voltage again builds up in preparation for the next spark. The repetition frequency is determined by the speed of rotation and the number of electrodes ; for example, a speed of rotation of 3,000 revolutions per minute or 50 per second and ten moving electrodes passing one fixed electrode gives a repetition

[5] F. S. Goucher, J. R. Haynes, W. A. Depp, E. J. Ryder, "Spark gap switches for radar," *Bell Syst. Tech. J.*, Vol. 25, 1946, pp. 563-602.

frequency of 500 cycles per second. Rotary gaps are not suitable for frequencies greater than about 1,000 cycles per second owing to the difficulty of separating the electrodes sufficiently rapidly to eliminate premature sparking while the network is being charged. Two types of construction are used. The first employs short rods mounted on an insulated disc which carries them between two fixed electrodes. The spark then jumps two short gaps in series. The second uses a metal disc to carry the moving electrodes, the disc being earthed by means of a carbon brush or collector ring. These are illustrated in Fig. 46.

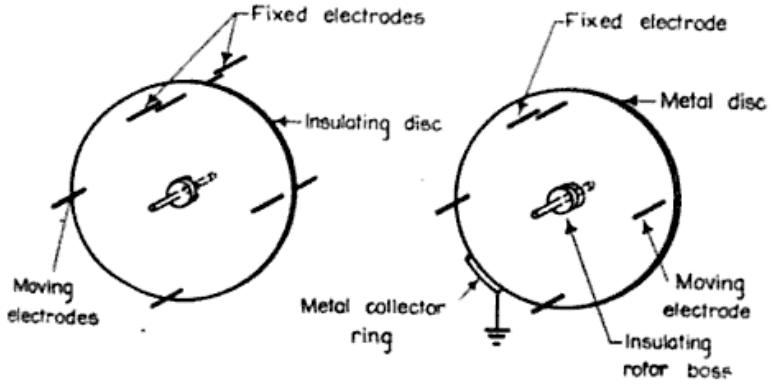


Figure 46.—Two types of rotary spark gap.

The operation of rotary gaps is satisfactory in the range from about 8 to 25 kilovolts. Large currents can flow, the only ill effects being wear of the electrodes. With tungsten electrodes and currents of a few hundred amperes the electrodes require adjustment every thousand hours. Breakdown of a gap is complete in a time of the order of 10^{-8} seconds after its initiation. The residual voltage across the gap is so small that it is difficult to measure and is probably less than 100 volts. Rotary gaps operate effectively in the open air but, unless of low power, are very noisy. It is therefore necessary to enclose them, provision being made for removing the products of the spark (ozone and oxides of nitrogen) from parts which may corrode.

The instant of sparking is irregular, a variation of the interval between pulses of ± 30 microseconds being usual. This makes it necessary to trigger the time base of a radar equipment from the modulator and precludes the use of simple rotary gaps in certain applications which require precise timing.

The ease of construction, simplicity of adjustment, and reliability of the rotary spark gap have contributed largely to the success of line type modulators.

Open Fixed Triggered Gaps. Triggered gaps were developed at the same time as rotary gaps with the object of eliminating moving parts and irregularity in firing. A common type illustrated in Fig. 47 consists of two spherical molybdenum electrodes with a tungsten electrode protruding through a hole

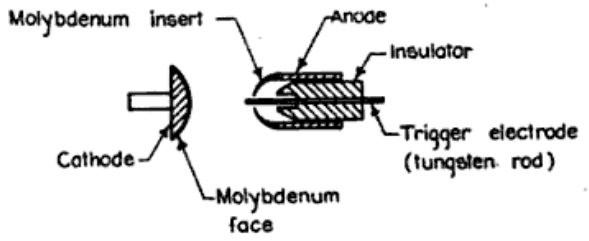


Figure 47.—Open fixed triggered gap.

in the anode. The spacing between the spheres is normally sufficient to prevent a spark occurring. At the desired moment a low power discharge is passed between the trigger electrode and the sphere through which it projects. The resulting ions and ultra violet light cause breakdown of the main gap. After discharge is complete the ionisation decays and the gap is again able to withstand the applied voltage. The switch illustrated is intended for a high power modulator and can handle frequencies from 100 to 2,000 cycles per second. Adjustments are required every 100 hours or so and the life is much less than for a rotary spark gap. The range of voltage for satisfactory operation is restricted but can be extended by blowing or pressurising the gap. A spark gap of this

type developed by Metropolitan Vickers Limited could switch 2 megawatts at 18 kilovolts at repetition frequencies up to 2,000 cycles per second. The switch fired within 0·1 microseconds of the application of the trigger pulse.

A typical trigger circuit is shown in Fig. 48. A pulse is applied to the grid of the tube, driving it positive so that a current builds up in the inductance L. The grid is then

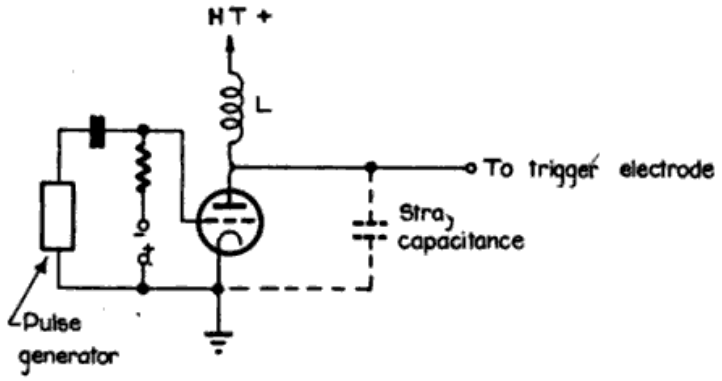


Figure 48.—Trigger circuit.

driven suddenly negative, cutting off the plate current and producing across the inductance a high voltage pulse which is applied to the trigger electrode.

Open gaps, both fixed and rotary, if used for airborne applications, must be housed in an air-tight enclosure to overcome the variations in atmospheric pressure encountered. This difficulty may be avoided by the use of an enclosed gap.[6]

Enclosed Fixed Triggered Gaps. There are two main types of enclosed fixed triggered gaps in general use. One type, (ii) in Plate VII (*b*) of which the CV85 is typical, is a modification of the open triggered gap and has the electrodes enclosed in a mixture of argon and oxygen. Time jitter of the discharge is of the order of 0·1 microseconds. The other type consists of two aluminium electrodes enclosed in a mixture of hydrogen and argon at a little less than atmospheric pressure. Because no

[6] J. D. Craggs, M. E. Haine, J. M. Meek, "Development of triggered spark gaps for high-power modulators," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 963-76.

trigger electrode is provided, the tubes must be used two or more in series with the trigger voltage applied at the common point of the tubes. Like the CV85, this tube, the 1B22 (i) in Plate VII (*b*) handles powers of about 100 kilowatts.

Thyratrons. The characteristics of a thyratron appear to make it ideally suited for application as a line type modulator switch. However, the high values and rapid rates of rise of currents proved beyond the capabilities of the types previously available. More recent work has led to the development of new forms which have satisfactory characteristics but which are difficult to manufacture. Of these, hydrogen thyratrons[7] appear the best and as low power tubes are used extensively. Hydrogen thyratrons have also been designed for power levels up to 15 megawatts. One tube, the 4C35, has a life of 900 hours at 8,500 peak volts and 90 amperes peak current with a 0·5 microsecond pulse recurring 2,000 times per second. The trigger voltage required is 150 volts rising at a minimum rate of 150 volts per microsecond. The time jitter is less than 0·04 microseconds using an AC heater supply and can be made undetectable by the provision of a DC heater supply.

4. Line Type Modulator Charging Circuits

At the cessation of a pulse, a pulse forming network is completely discharged and must be recharged before the commencement of the next pulse. Since the charging period is long compared with the pulse duration, the pulse forming network behaves like a simple capacitor as far as the charging circuit is concerned. Since this capacitor stores the energy for the pulse, its capacitance is given by

$$\tfrac{1}{2}CV_o^2 = \frac{\delta V^2}{R}$$

where C = line capacitance (farads), V = pulse voltage,

[7] R. H. Wittenberg, "Thyratrons in radar modulator service," *R.C.A. Rev.*, Vol. 10, 1949, pp. 116-33; and H. de B. Knight, L. Herbert, "Development of mercury-vapour thyratrons for radar modulator service," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 949-62.

V_o = voltage to which network is charged, R = load resistance (ohms), δ = pulse duration (seconds).

The usual network is equivalent to a simple transmission line so that $V/V_o = \frac{1}{2}$ and $C = \delta/2R$. The average power which must be supplied to the pulse forming network is $\frac{1}{2}f_r C V_o^2$ where f_r is the repetition frequency.

If the pulse repetition frequency is different from the AC mains frequency, the pulse forming network can be charged from a conventional rectifier circuit with filter capacitors. Such a method is referred to as DC charging. If the repetition and mains frequencies are the same or simply related (e.g. half or twice) the filter circuits and often the rectifier may be dispensed with. This is called AC charging. In all cases the components connecting the power supply and the network must be capable of isolating the one from the other during and immediately after the pulse.

DC Charging Circuits

Resistance Charging. The simplest method of charging a pulse forming network is through a resistance connected in place of the inductance in Fig. 49. The efficiency of the system

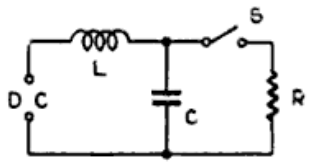


Figure 49.—DC inductance charging—basic circuit.

cannot exceed 50 per cent so it is seldom used except for low power modulators or for experimental purposes.

Inductance Charging. Fig. 49 shows the basic circuit for DC inductance charging, an arrangement which does not suffer from the intrinsic inefficiency of resistance charging. There are three main cases to consider: resonant charging, linear charging and diode charging. In all cases the network is

charged to a voltage equal to twice the DC supply voltage if losses are neglected. In practice the voltage step-up is about 1·9 times.

Resonant charging refers to the case in which the resonant frequency of L and C in the circuit is equal to half the repetition frequency f_r so that the charging current I_c and network voltage V_c pass through one half cycle between pulses. Fig. 50 shows the current and voltage waveforms of the charging cycle.

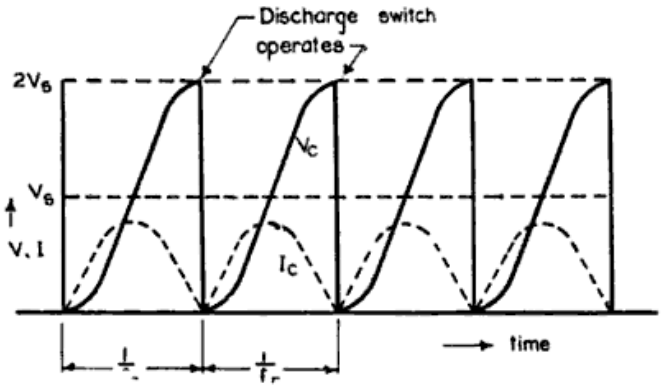


Figure 50.—Voltage and current waveforms in DC " resonant charging."

Since the network voltage remains almost constant for an appreciable period about the time the switch operates, this type of charging wave is suitable for use in modulators where the switch has a time jitter, such as a rotary spark gap. No over-voltage occurs if the switch fails to operate.

If the resonant frequency f is low compared with half the repetition frequency f_r, so called " linear charging " results. The large inductance used tends to make the charging current constant which in turn charges the network approximately linearly. Fig. 51 shows typical waveforms obtained with linear charging. This method allows the pulse repetition rate to be varied over a wide range without changing the value of the inductance. Satisfactory results are obtained if the inductance is made about four times that required for resonant charging at the lowest repetition frequency. If the switch should fail to operate, the voltage rises in the manner indicated

by the dotted line in Fig. 51. If this is excessive the network can be protected with a spark gap. A thyratron or fixed gap should be used for the switch as irregular firing results in fluctuations of the output pulse voltage.

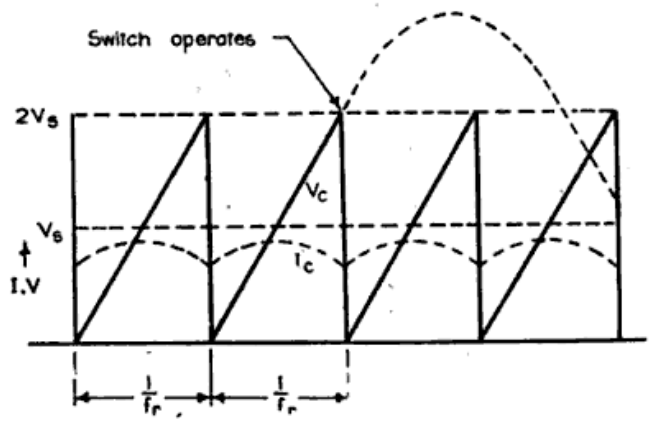


Figure 51.—Voltage and current waveforms in DC "linear charging."

So called "diode charging" results from the use of a diode in series with an inductance less than that required for resonant charging. The network charges in half the resonant period of L and C, which is less than the interval between pulses,

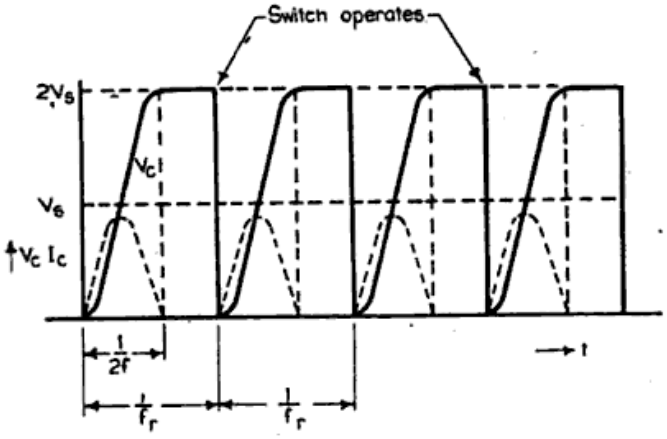


Figure 52.—Circuit and current voltage waveforms in DC "diode charging."

and the diode then prevents the discharge of the network till the switch operates. The voltage and current waveform are shown in Fig. 52. This circuit has the advantage that it can

handle a wide range of repetition frequencies and, as the voltage is steady prior to discharge, the irregular firing of a rotary spark gap does not produce pulses of irregular amplitude. It suffers from the disadvantage of introducing another high voltage component.

The inductances used in DC charging can be quite readily designed from a knowledge of the incremental permeability of the core material for the currents flowing. The iron core is best utilised without an air gap and at a value of magnetising force of about 4,000 amperes per metre (50 oersteds) at the current peak. The inductance values range from one to several hundred henries depending on the requirements.

DC Power Supplies for Line Type Modulators

Modulators use conventional rectifier circuits to convert AC to DC power, but the problem of the exact calculation of "ripple voltage," or the variation in the voltage to which the pulse forming network is charged, is often difficult. Design information obtained experimentally is given in Fig. 53 for a simple condenser input filter with inductance charging of the network. As given it applies to a half wave rectifier; for a full wave rectifier the supply frequency must be doubled.

If the supply frequency is much less than the repetition frequency, the filter capacitor must be sufficiently large to charge the network several times without its voltage falling unduly. If the supply frequency is higher than the repetition frequency a great reduction in the size of the filter capacitor is possible, since it will receive one or more current impulses from the rectifier during each charging interval of the network. For example, if a 50 cycle full wave rectifier charges a network 500 times per second, a simple calculation shows that the capacitance must be 1,000 times that of the network if the ripple voltage is to be 1 per cent. However, if the supply frequency is about 500 cycles per second, it need only equal

that of the network. The economy resulting from the use of a high supply frequency applies also to the rectifier transformer and other parts of the radar equipment.

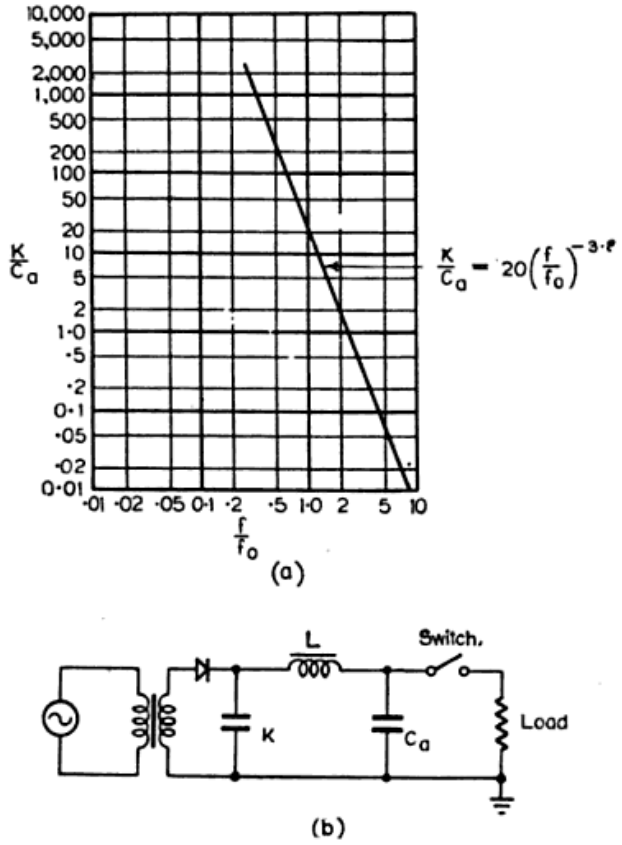


Figure 53.—Ratio of filter capacitance to network capacitance for 1% ripple.

K = filter capacitance
C_a = network capacitance
f = supply frequency
f_0 = repetition frequency.

(*a*) Experimental relation between C and K. (*b*) Basic circuit.

Charging from Low Voltage DC Source

If low voltage DC only is available, say from a generator or battery, the circuit in Fig. 54 enables a pulse forming network

to be charged to any required voltage. The arrangement is essentially the same as for DC inductance charging, except that the inductance used is the transformer leakage inductance. In addition the voltage step-up at the pulse

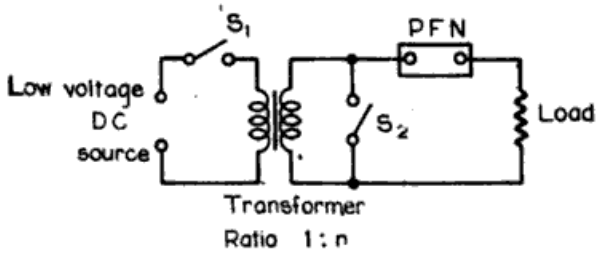


Figure 54.—Basic circuit for DC charging from low voltage source.

forming network is increased by the transformer turns ratio. The switch S_1 is closed during the charging period and opens just before switch S_2 closes to discharge the network into the load. S_1 closes again after the pulse and the cycle is repeated. Switch S_1 is a low voltage and switch S_2 is a high voltage switch. As they must be accurately triggered, thyratrons are the most satisfactory switches for this purpose.

AC Charging Circuits[8]

In the circuits described as AC charging circuits the filter capacitors are omitted, the pulse network being charged by a transformer through a suitable isolating device. The power supply frequency must then be equal to, or simply related to, the pulse repetition frequency. If a repetition frequency of some hundreds of cycles per second is required, it is usual to include a high frequency alternator in the modulator. This alternator may supply power to the whole radar set thus economising in the size of other components. If a rotary spark gap is used, it can be mounted on the alternator

[8] K. J. R. Wilkinson, "Some Developments in high-power modulators for radar," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 1090-112.

shaft with provision for correctly phasing the instant of discharge of the network with respect to the output voltage of the alternator.

Fig. 55 shows a practical arrangement in which a diode is used to isolate the pulse forming network at the time of discharge. The figure gives waveforms and indicates the switch

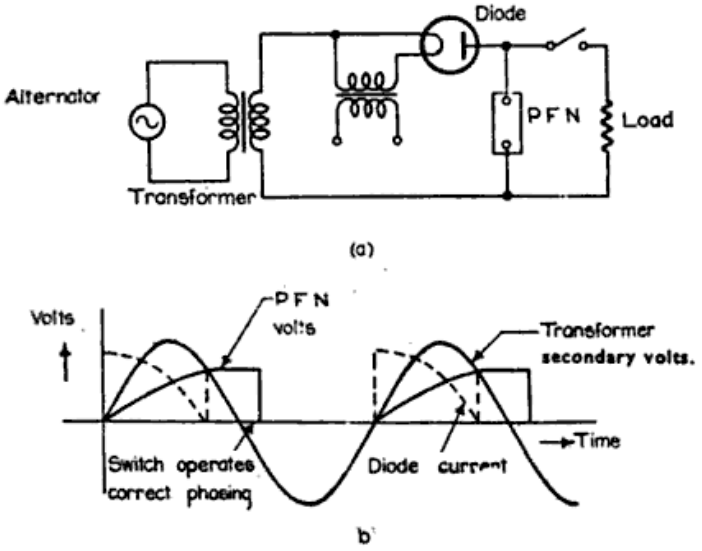


Figure 55.—(*a*) AC diode charging circuit. (*b*) Waveforms.

phasing which keeps the diode inverse voltage as low as possible. This circuit is simple to adjust but since power is supplied to the network during a quarter of a cycle only the root-mean-square current is unduly great and the transformer is not used to the best advantage.

The diode and its filament transformer are high voltage components which can be dispensed with, as in the more efficient arrangement of Fig. 56 (*a*) which uses an inductance to isolate the network from the transformer. An equivalent circuit is shown in Fig. 56 (*b*) in which all quantities are referred to the transformer secondary terminals. The inductance L is the equivalent inductance including contributions from leakage in the transformer and the generator. The resistance is

usually neglected. The performance may be described in terms of the ratio, k, of the resonant frequency of L and C to the supply frequency.

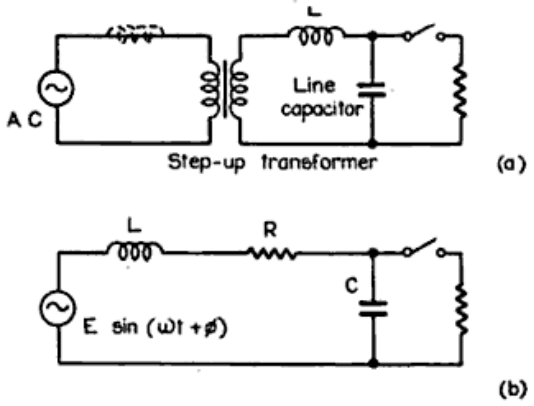


Figure 56.—AC inductance charging circuits. (*a*) Actual circuit. (*b*) Equivalent circuit.

If these frequencies are equal the condition is known as "AC resonant charging." In this case the pulse forming network is charged to a voltage π times the equivalent electromotive force. The switch operates as the electromotive force passes through zero, and the network is completely recharged in one cycle, the process being repeated every cycle. The pulse

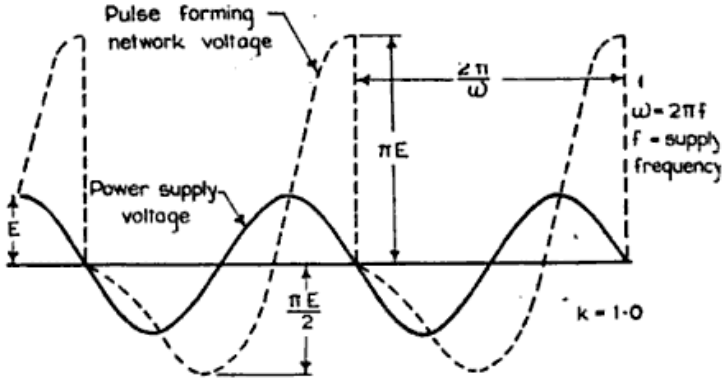


Figure 57.—AC resonant charging voltage waveforms.

network polarity depends on the zero of the applied electromotive force chosen. The transformer secondary carries a direct component of current equal to approximately one third

of the root-mean-square current. Voltage waveforms are shown in Fig. 57. It is obvious that if the switch fails to operate large voltages will tend to build up but this may be prevented by fitting protective spark gaps.

In general the voltage to which the pulse network is charged will depend on k in the manner shown in Fig. 58 provided the switch fires at the optimum phase. When $k = 1{\cdot}4$ the greatest

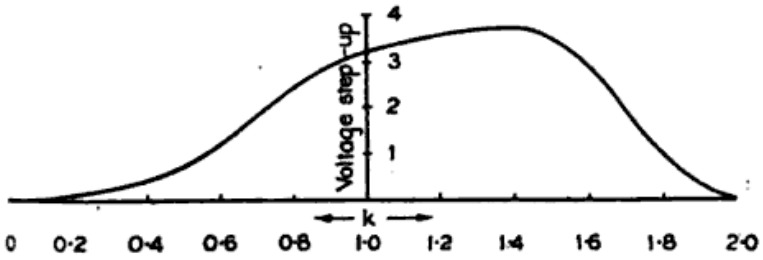


Figure 58.—AC charging. Voltage step-up as a function of k.

step-up occurs, being 3·7 as against π for the resonant case. To obtain this step-up ratio the network should be discharged 114 electrical degrees past the zero of the applied electromotive force. Waveforms for this case are shown in Fig. 59.

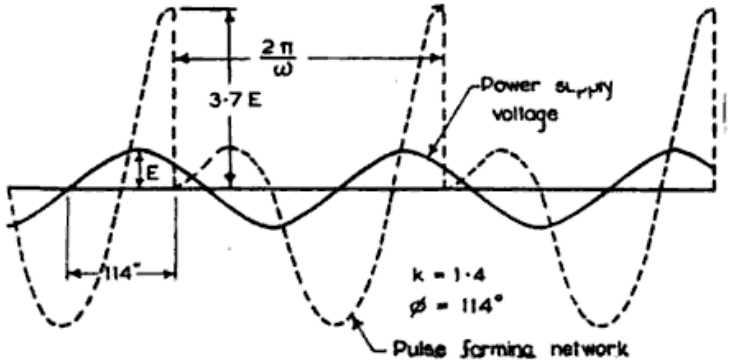


Figure 59.—AC charging waveforms for $k = 1{\cdot}4$.

A feature of AC charging is the remarkable economy of components it permits. The accompanying disadvantage is lack of flexibility in pulse length and repetition frequency, which is serious in the case of laboratory equipment, but is often of no importance for service use. Fig. 60 gives a

summary of the main characteristics of the modulator charging circuits discussed above.

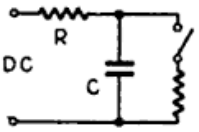
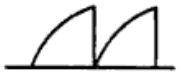
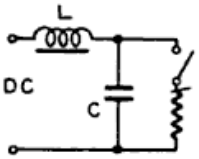
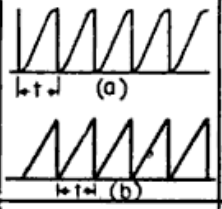
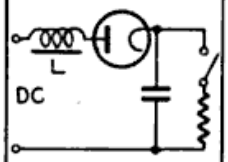
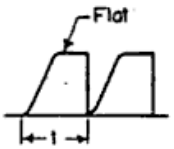
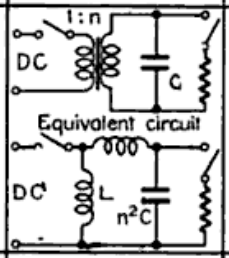
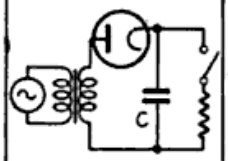
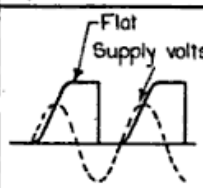
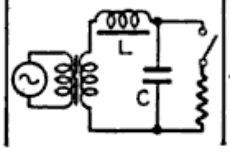
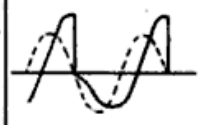
Title	Circuit	Typical Waveform	Remarks
DC Resistance Charging			Efficiency low
DC Inductance Charging (a) Resonant (b) Linear			(a) $\pi\sqrt{LC} \doteq t$ (b) $\pi\sqrt{LC} >> t$
DC Inductance Diode Charging			$\pi\sqrt{LC} < t$
DC Transformation Charging			$n\pi\sqrt{LC} < t$ L = Transformer leakage inductance Operated from low voltage DC source
AC Diode Charging			
AC Resonant Charging			L may be the transformer leakage inductance

Figure 60.—Summary of characteristics of modulator charging circuits.

5. Pulse Transformers

As noted earlier, the introduction of pulse transformers in conjunction with line type modulators permitted a great saving in bulk and weight and also made it possible to separate the modulator from the transmitter.

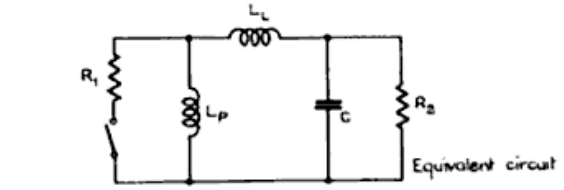
Pulse transformers may be required to transmit rectangular pulses of the order of 1 microsecond duration and 1 megawatt peak power without significant distortion. A standard input impedance is 50 ohms while the output impedance when feeding a magnetron is about 800 ohms. For a power output of 1 megawatt the input voltage is therefore 7000 and the output 28000 volts. These requirements introduce serious design difficulties both in the reproduction of such short pulses and in the maintenance of adequate insulation.

The problem can be more clearly formulated if we consider the approximate equivalent circuit of a transformer given in Fig. 61. In this circuit all quantities are referred to the primary in the conventional way. The leakage inductance is represented by L_L, the primary inductance by L_p, and C represents the equivalent capacitance across the secondary due to winding capacitances and strays in both the transformer and load. The losses and the more complex stray capacitance distribution which exist in practice are neglected.

If the distortion is not great, the effects due to leakage inductance and stray capacitance may be separated from those due to the primary inductance. The former distort the rise and fall of the pulse causing them to take the form of damped oscillations associated with the circuit R_1, R_2, L_L and C, which is obtained by omitting L_p from Fig. 61.

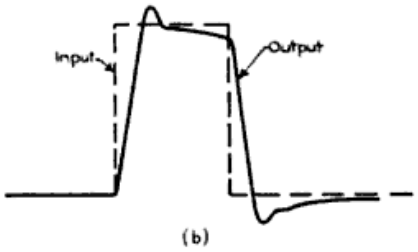
The primary inductance draws an increasing current through R_1 during the pulse and this causes a fall in the top of the pulse and an overshoot at the end as indicated in Fig. 61. Both types of distortion may be determined quantitatively if the circuit elements are known. To reduce distortion it is necessary to achieve a combination of high primary inductance, low leakage inductance and low winding capacitance. These conflicting requirements have been met by paying close

attention to the quality of the magnetic circuit so that physical size can be reduced to a minimum while maintaining the necessary primary inductance. This leads to a reduction of capacitance and leakage inductance but increases the insulation problem.



R_1 Characteristic impedance of pulse forming network
L_P Magnetising inductance
L_L Leakage inductance
C Effective shunt capacitance
R_2 Load impedance
(All quantities referred to are primary or secondary)

(a)



(b)

(b)

Figure 61.—(*a*) Equivalent circuit of pulse transformer. (*b*) Input and output waveforms.

If a core of normal construction were used in a pulse transformer, skin effect would severely reduce its effective permeability. This difficulty is overcome by forming the core

from exceptionally thin laminations and by using a core material having a high resistivity which tends to increase the depth of penetration at high frequencies.

In designing a pulse transformer[9] the primary, secondary and leakage inductances are estimated in the manner used for normal transformers. The effective winding capacitance is determined by assuming a voltage distribution on the windings, estimating the capacitative energy associated with significant sections of them, and equating the sum of these energies to $\frac{1}{2}CV^2$ in the equivalent circuit.

Design Procedure

Data required when designing a pulse transformer include pulse duration, primary and secondary pulse voltages and currents, and the admissible peak value of magnetising current (of the order of 10 per cent of the desired pulse current). The design steps follow broadly the sequence set out below :

(*a*) The best available core material is selected, and suitable peak values of magnetic induction B and magnetic intensity H are chosen using curves obtained under actual pulse conditions.

(*b*) The core shape is chosen empirically after the core volume is calculated from equation (1) which is obtained by equating the magnetic energy of the core to the energy delivered to it.

$$A\,l \times B\,H = V_p \delta I_m \qquad (1)$$

where A = cross sectional area of core (square metres)
l = core length (metres)
B = peak magnetic induction (webers per square metre)
H = peak magnetic field intensity (amperes per metre)

[9] W. S. Melville, "Theory and design of high-power pulse transformers," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 1063-80; N. F. Moody, "Low-power pulse transformers," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 311-4; R. Lee, *Electric Transformers and Circuits*, John Wiley and Sons, New York, 1947, pp. 219-57.

V_p = primary pulse voltage
δ = pulse duration (seconds)
I_m = peak magnetising current (amperes).

(*c*) The number of turns required on the primary and secondary windings is determined from equation (2) and wire sizes are allotted to give suitable current densities.

$$V_p\delta = N_pAB \tag{2}$$

where N_p is the number of primary turns.

(*d*) An arrangement of the windings is chosen to fit the core window and to give adequate insulation. The different portions of the windings are interleaved so as to keep leakage inductances and winding capacitances low.

(*e*) If considered necessary, leakage inductance and winding capacitances can be estimated by methods outlined above and their effect on the output waveform investigated.

Construction

The Core. Core materials combine high resistivity with high permeability at high values of magnetic induction. They are made in laminations about 0·002 inches thick. Typical B-H curves for complete cores when 1 microsecond pulses are used are given in Fig. 62.

In order to obtain a magnetic circuit of the lowest reluctance special precautions must be taken to minimise the effect of joints in the core. Some designs use continuous cores, but special coil winding methods must then be adopted. This difficulty is overcome in one interesting construction in which a continuous core is first wound from strip around a rectangular mandrel and then cut to permit the insertion of coils

wound in the normal way. The core is rejoined after the faces of the joints have been ground flat and etched to remove inter-laminar short circuits.

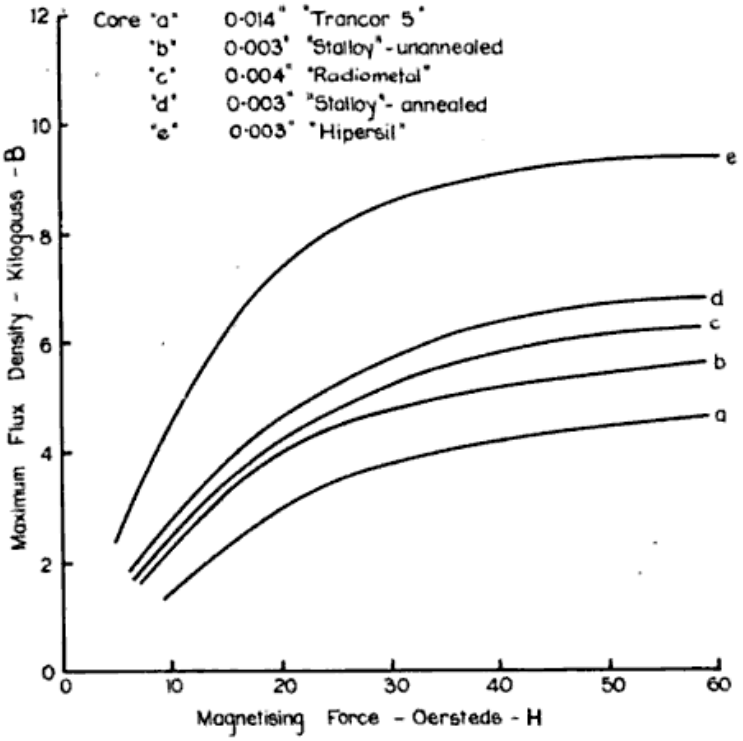


Figure 62.—B-H curves for pulse transformer core materials.

Windings. The winding configuration is determined by the conflicting requirements of low leakage inductance which requires windings in close proximity, and low capacity and high insulation which require the separation of points differing appreciably in potential. Separated single layer windings are used. As most of the heat dissipated in pulse transformers is due to core losses, very high current densities are used in the windings viz. 5000 amperes per square inch for continuous loads and 3000 amperes per square inch for pulse currents.

A special feature of pulse transformers designed for driving magnetrons is the use of a bifilar secondary winding which

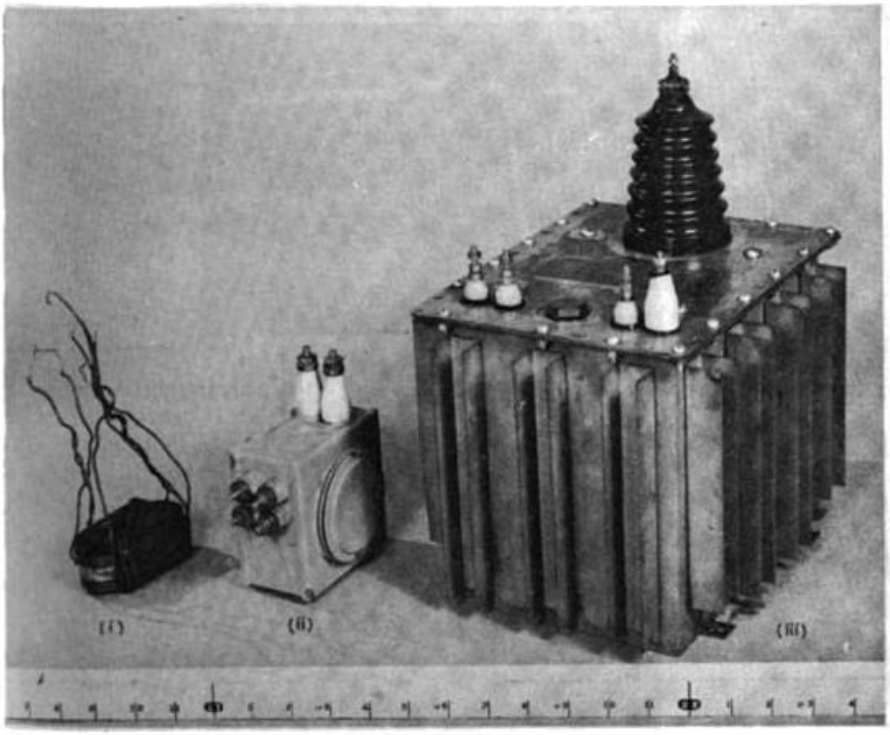
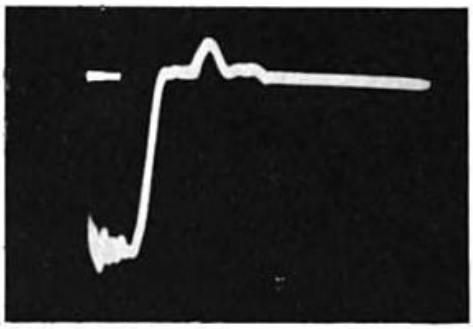
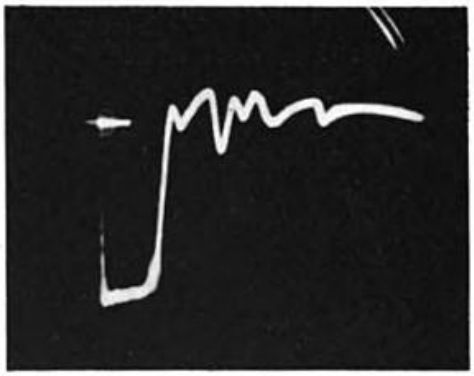


PLATE VIII.—Output pulse transformers.

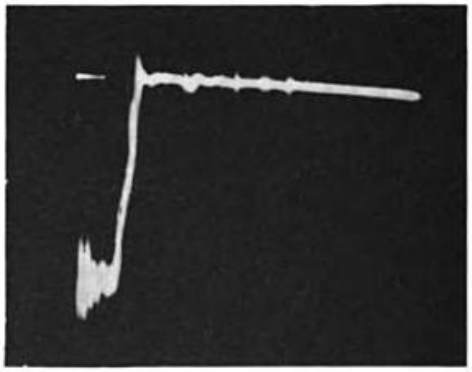
facing p. 106]



(*a*) Voltage waveform—modulator discharging into a non-inductive resistor.



(*b*) Voltage waveform—magnetron load.



(*c*) Current waveform—magnetron load.

PLATE IX.—Waveforms for line type modulator (see p. 109).

supplies the magnetron heater current, thus eliminating a separate heater transformer of high insulation.

Insulation problems encountered in pulse transformers are extraordinary and they have been overcome only by close attention to refinements of impregnating techniques. The best transformer oil is used and exceptionally thorough vacuum impregnation and hermetic sealing are essential. Kraft paper is used almost universally for insulation between winding layers and when oil immersed may be stressed up to 300 kilovolts per inch. The oil itself may be stressed up to 500 kilovolts per inch while over the surface of the paper the stress should not exceed 30 kilovolts per inch. Very high voltage stresses between turns are encountered, values of up to 500 volts per turn being quite common. Plate VIII shows photographs of three typical pulse transformers. Plate VIII (i) shows an uncased transformer with a "hipersil" core wound on a mandrel and cut as described on p. 104. Plate VIII (ii) shows a 125 kilowatt 1 microsecond pulse transformer designed for airborne use. The larger transformer shown in Plate VIII (iii) is a laboratory model for 6 microsecond pulses with an output of 2 megawatts 40 kilovolts.

6. Typical Modulator Circuits

High Vacuum Tube Modulator

The circuit of a typical high vacuum tube modulator designed for airborne microwave equipment is shown in Fig. 63. It generates negative pulses of 11 kilovolts and 10 to 15 amperes and of 1 microsecond duration at a repetition frequency which may be varied between 400 and 2000 cycles per second. The pulses are formed at a low power level by discharging a pulse forming network through a small thyratron V_2 which is triggered from an external source. The positive pulse so obtained is amplified by the twin tetrode V_3 which is cathode coupled to the high power switching tubes V_6 and V_7. The whole pulse forming circuit follows the potential of the cathode of V_3, so that this is an example of the so-called " boot-strap "

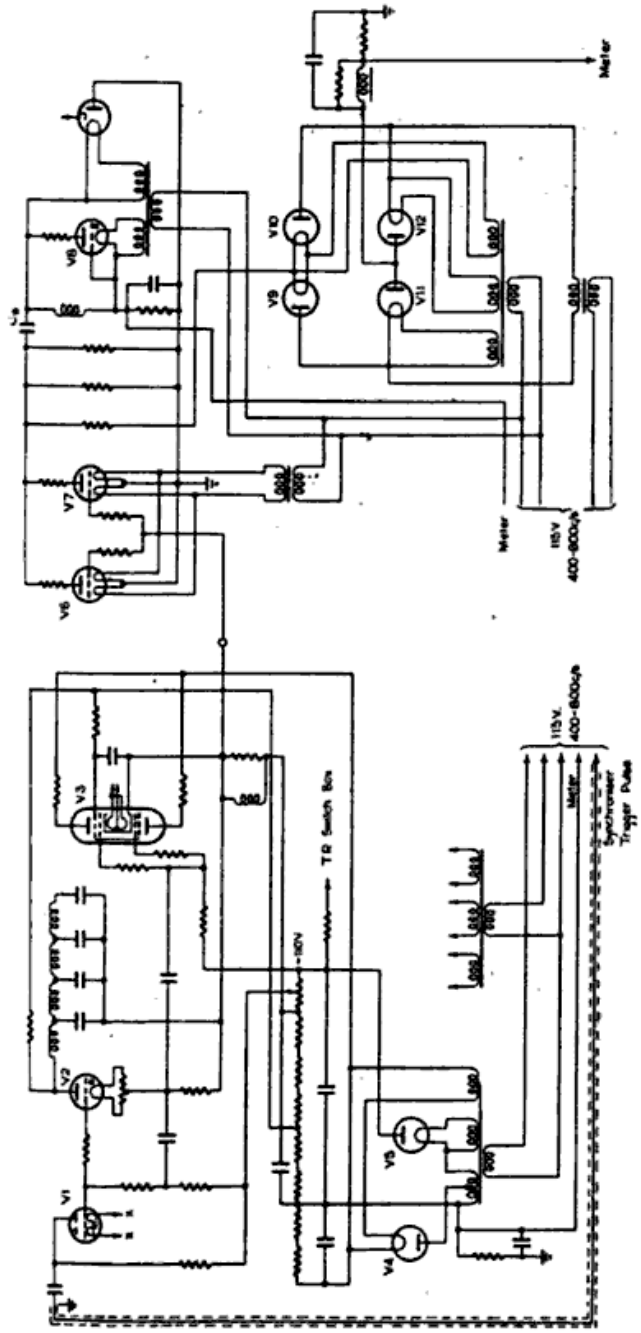


Figure 63.—High vacuum tube modulator circuit developed by Radiation Laboratory, Massachusetts Institute of Technology.

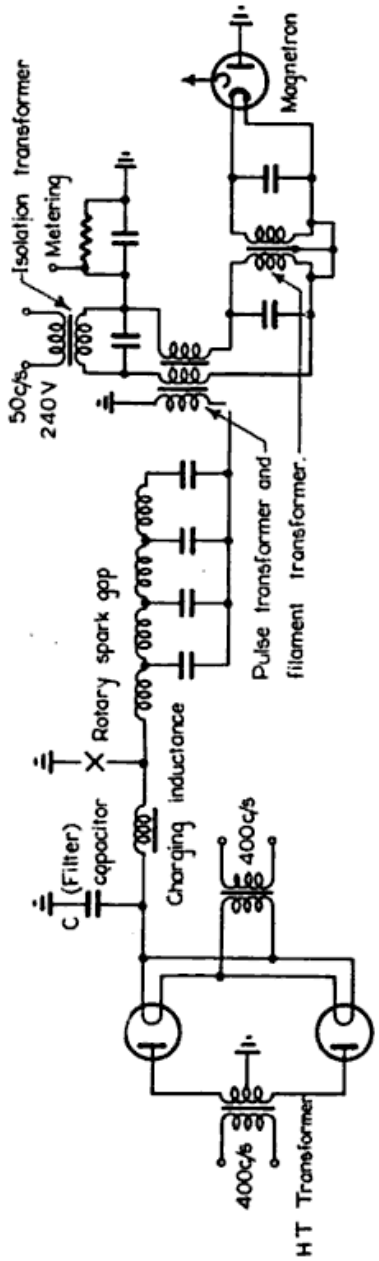


Figure 64.—Line type modulator—resonant charging circuit.

circuit. The system operates on 115 volts 400-800 cycles per second, requires an input of 1 kilowatt, and has an overall efficiency of from 10 to 30 per cent. The weight is about 60 pounds.

Line Type Modulator with Rotary Spark Gap

Fig. 64 shows the circuit of a line type modulator designed for laboratory test purposes. The unit is mounted on a 400 cycles per second alternator set. It uses DC resonance charging, a rotary spark gap and a pulse transformer. It delivers a pulse of 40,000 volts at 50 amperes or a peak power output of 2 megawatts. Two combinations of pulse length and repetition frequency are available, 1 microsecond at 500 cycles per second, and 5 microseconds at 250 cycles per second.

The output waveforms obtained from this modulator are typicalof radar practice and are shown in Plate IX. Plate IX (*a*) shows the voltage waveform when power is delivered to an 800 ohm non-inductive resistor. The oscillations on top of the pulse are characteristic. The second oscillogram, Plate IX (*b*), shows the voltage pulse with a magnetron as load. The flat top of this pulse illustrates the effect of the constant voltage characteristic of the magnetron. The third oscillogram, Plate IX (*c*), that of magnetron current, shows oscillations which are a little greater than those with a resistive load.

CHAPTER VI

MICROWAVE TRANSMISSION AND CAVITY RESONATOR THEORY

1. Microwave Transmission Lines

THE function of a transmission line is to transfer electromagnetic energy from one point to another with the least loss. The conventional transmission line of two conductors is well known as a means of doing this. Either the conductors are external to each other as in the case of the twin wire line or one completely surrounds the other as for the coaxial transmission line. In the former case the fields spread out into space and loss of energy results at short wavelengths. A coaxial transmission line however confines the field within the outer cylinder and therefore is not subject to attenuation by radiation. In the region of decimetre wavelengths twin wire lines become unusable. The transmission lines used at these and shorter wavelengths are the object of our attention here.

Well into the microwave region the coaxial line continues to be used until accumulating disadvantages limit its further

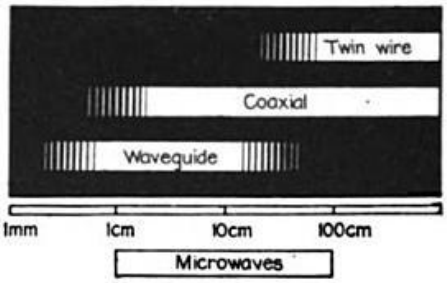


Figure 65.—Field of use of various transmission systems.

application. In addition to the two conventional types of transmission line there is the hollow pipe or waveguide. Unlike the former which find use from the longest wavelengths down, the waveguide does not become practical until short wavelengths are reached. Fig. 65 illustrates the relation between the regions in which the various lines are used. It

will be seen that the twin line is of no importance in microwave transmission but that coaxial and waveguide transmission are both met over nearly the whole region.

The word *microwave* is roughly synonymous with centimetre-and-decimetre waves but its real use is to designate certain techniques—consequently definite wavelength limits cannot be set to the microwave region. The techniques symptomatic of microwaves involve the use of completely enclosed apparatus such as waveguides and electromagnetic cavities.

2. Waveguides

A waveguide is simply a hollow conducting pipe through which electromagnetic waves will travel. It differs from the conventional transmission line in having no return conductor and therefore its operation cannot be visualised in terms of current and voltage waves—though the idea of impedance can be retained. To deal with waveguides it is necessary to move the emphasis from voltages and currents to the electric and magnetic fields confined within the guide. Virtually all the field is restricted to the dielectric medium filling the guide; for alternating fields cannot exist within a perfect conductor and penetrate only slightly into practical conductors at radio frequencies.

Many different field patterns can be propagated along a waveguide or transmission line but with a two-conductor line emphasis falls on one only of these modes, the conventional mode for which the electric and magnetic fields are purely transverse. The existence of possible higher modes of propagation on a twin wire or coaxial line may usually be safely forgotten for a reason which appears below. However in a hollow pipe the conventional or Transverse Electro-Magnetic (TEM) mode cannot exist and attention must be focused on the higher modes of propagation. They are conveniently classed as transverse electric and transverse magnetic according to whether the electric or magnetic field remains purely transverse. Hence their distinguishing feature is the presence of a component of field in the direction of propagation. If longitudinal components of both field types are present the

pattern is regarded as a linear combination of transverse electric and transverse magnetic modes and each is treated separately. In general the phase velocities of the components will not be equal and the pattern will change as it travels.

Figs. 66 and 67 illustrate fields in coaxial line and circular waveguide. In the former the field components are completely transverse while in the waveguide, although the magnetic

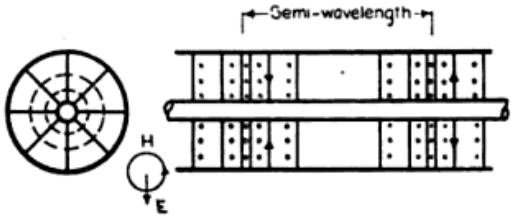


Figure 66.—Conventional mode in coaxial transmission line.

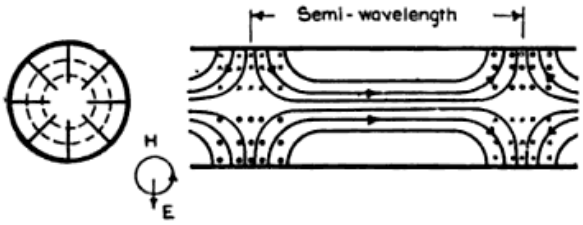


Figure 67.—TM_{01} mode in circular waveguide.

field is similar, the electric lines turn out of the transverse plane and make their way back to the outer wall, having no central conductor on which to terminate.

The property of higher modes which distinguishes them from the conventional mode, is the possession of critical wavelengths. Thus although waves of any frequency can negotiate a two-conductor transmission line, the absence of the conventional mode from a waveguide divides the spectrum into transmitted and non-transmitted frequencies. For example, electromagnetic waves at optical frequencies will pass down a hollow pipe of moderate size whereas waves at power frequencies will not. These extreme cases are well away from the critical wavelength. Microwave transmission through a waveguide is carried on at a wavelength just shorter than the critical

wavelength. Another way of stating this is to say that the waveguide is made just large enough to transmit the wavelength in question.

Each possible field pattern in the waveguide has its own critical wavelength which must not be exceeded if there is to be propagation. A waveguide is usually designed so that it is large enough to transmit in the mode with the longest critical wavelength (the dominant mode) but not large enough to transmit in the next higher mode. If in a particular transverse section of a guide a field pattern is applied which cannot propagate, then the pattern is attenuated exponentially along the guide at a rate depending on how close it is to its critical wavelength. In general the higher the order of a mode the more rapidly it is damped out. Fig. 68 shows how the field

Figure 68.—TM_{01} mode in pipe too small to allow propagation.

in a circular waveguide falls off when a circular magnetic field is applied at the mouth. The mode is the same as that which in Fig. 67 is shown travelling down a pipe of larger size.

It can now be seen why the existence of higher modes is not of importance in conventional transmission lines—for at wavelengths longer than the critical wavelength no energy can be transferred save by the conventional mode. As the longest critical wavelength in a coaxial line cannot exceed the circumference of the outer conductor only at the shortest wavelengths do new considerations arise. The part played by attenuated modes in coaxial lines is one of adjusting boundary conditions. For example the field impressed on the mouth of a coaxial line may be neither circularly symmetrical nor inversely proportional to the radius, yet a short distance along the line the field will be quite pure. This is accounted for by analysing the applied field into an infinite series of orthogonal patterns all of which save one are attenuated to negligible

proportions in a distance of the order of a diameter. That pattern which alone can propagate therefore exists in complete purity save in the immediate neighbourhood of discontinuities.

In a waveguide whose surface is perfectly conducting a propagated wave suffers no attenuation, but in a waveguide of finite conductivity the current flow in the walls results in the development of heat which appears at the expense of the energy of the field. As the amount of heat liberated per unit distance at any section is proportional to the power crossing the section, the wave is attenuated exponentially with distance just as waves on conventional transmission lines. This sort of attenuation, which is due to conductor losses, represents a real loss of energy whereas the attenuation in a pipe too small to allow propagation is reactive.

In choosing between the waveguide and coaxial line for a particular application four factors are considered. These are dielectric loss, conductor loss, power handling capacity and mechanical convenience. Without considering details a tendency may be observed for waveguide to replace coaxial line as the wavelength shortens, first for the transmission of high powers and later for local oscillator and signal powers. Thus a decimetre-wave radar may have large air-filled and small polythene-filled coaxial lines for transmitter and receiver respectively, a set operating on short centimetre waves may have waveguide in each case, whereas at frequencies between the extremes it is common to find the transmitter power carried by waveguide and the low powers by flexible polythene cable.

We have seen that there are numerous modes of transmission through waveguides each with its own critical wavelength and phase velocity and that at frequencies higher than the critical frequency there is a damped travelling wave, whilst at lower frequencies there is an attenuated standing wave. To illustrate the derivation of these and other results we consider in detail the rectangular waveguide.

3. Rectangular Waveguide

Solution of Maxwell's Equations

The following theory applies to transverse electric modes in rectangular waveguides. In a passive medium characterised by permeability μ, permittivity ϵ and conductivity g, Maxwell's equations for harmonic electric and magnetic fields are

$$\text{curl } \mathbf{E} = -j\omega\mu\mathbf{H} \tag{1}$$

$$\text{curl } \mathbf{H} = (g + j\omega\epsilon)\mathbf{E} \tag{2}$$

in rationalised m.k.s. units.[1] A factor $\exp(j\omega t)$ is omitted from all field expressions. To specify a medium we may take instead of μ, ϵ and g the derived constants η and σ such that

$$\eta = \sqrt{\frac{j\omega\mu}{g + j\omega\epsilon}} = \mathsf{R} + j\mathsf{X} \tag{3}$$

and

$$\sigma = \sqrt{j\omega\mu(g + j\omega\epsilon)}, \tag{4}$$

where η is the intrinsic impedance of the medium, σ is the intrinsic propagation constant and R and X are the intrinsic resistance and reactance. The propagation constant of a plane wave in the medium is the same as σ whilst η is the ratio of the electric to the magnetic field of a plane wave in the medium. The same advantages are gained by using these derived constants as are gained in transmission line theory by working with characteristic impedance and propagation constant instead of the distributed parameters L, C, R, G.

We desire to find solutions of equations (1) and (2) inside a rectangular pipe extending in the z-direction and bounded by the perfectly conducting planes $x = 0, a$ and $y = 0, b$. On these planes there must be no tangential component of electric

[1] The rationalised metre-kilogram-second system of units is used throughout this book. For information on this system of units see J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill, New York, 1941. The quantities appearing in Maxwell's equations are measured in the following units :—$\mathbf{E}$ (electric field strength), volts per metre ; $\mathbf{H}$ (magnetic field strength) amperes per metre ; g (conductivity), mhos per metre ; μ (permeability), henries per metre ; ϵ (permittivity), farads per metre. The permeability of free space (μ_0) is $4\pi \times 10^{-7}$ henries per metre ; the permittivity of free space (ϵ_0) is $8{\cdot}85 \times 10^{-12}$ farads per metre and the intrinsic impedance of free space (η_0) is given by $(\mu_0/\epsilon_0)^{\frac{1}{2}} = 120\pi = 377$ ohms.

field and no normal component of magnetic field. In Cartesian coordinates the equations become

$$\begin{aligned}
\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -j\omega\mu H_x \\
\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -j\omega\mu H_y \\
\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -j\omega\mu H_z \\
\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= (g + j\omega\epsilon) E_x \\
\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= (g + j\omega\epsilon) E_y \\
\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= (g + j\omega\epsilon) E_z.
\end{aligned} \tag{5}$$

We shall now assume that a solution exists such that the electric vector is everywhere parallel to the y-direction i.e. that $E_x = E_z = 0$. By assuming $E_z = 0$ we are restricting the solutions to the transverse electric class. By the further assumption that $E_x = 0$ we shall be left with a special set of transverse electric waves. The equations now become,

$$\begin{aligned}
H_x &= \frac{1}{j\omega\mu}\frac{\partial E_y}{\partial z} \\
H_y &= 0 \\
H_z &= -\frac{1}{j\omega\mu}\frac{\partial E_y}{\partial x} \\
\frac{\partial H_z}{\partial y} &= \frac{\partial H_x}{\partial y} = 0 \\
\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= (g + j\omega\epsilon) E_y.
\end{aligned} \tag{6}$$

Eliminating **H**,

$$\frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial z^2} = \sigma^2 E_y \tag{7}$$

where

$$\sigma^2 = j\omega\mu\,(g + j\omega\epsilon). \tag{8}$$

From this partial differential equation for the single field component E_y we obtain the solution by separation of variables by substituting $E_y = X(x)Z(z)$. The resulting total differential equations for the sought functions X and Z give

$$X(x) = A \sin hx + B \cos hx \tag{9}$$

$$Z(z) = C \exp(\Gamma z) + D \exp(-\Gamma z) \tag{10}$$

where

$$h^2 = \Gamma^2 - \sigma^2. \tag{11}$$

To determine the integration constants we note that as E_y must vanish for $x = 0, a$

$$X(x) = A \sin \frac{n\pi x}{a},\ n = 1, 2, \ldots \tag{12}$$

and as the terms of Z evidently represent waves travelling in opposite directions we may take $C = 0$ and consider only waves travelling in the positive z-direction. Then if E is the maximum value of the electric field,

$$E_y = E \sin \frac{n\pi x}{a} \exp(-\Gamma z) \tag{13}$$

where

$$\Gamma = \sqrt{\left(\frac{n\pi}{a}\right)^2 + \sigma^2}. \tag{14}$$

The propagation constant Γ is equal to $\alpha + j\beta$ where α is the attenuation constant and β the phase constant.

For the remaining field components we have from (6)

$$H_z = -\frac{n\pi E}{j\omega\mu a} \cos \frac{n\pi x}{a} \exp(-\Gamma x) \tag{15}$$

$$H_x = -\frac{\Gamma E}{j\omega\mu} \sin \frac{n\pi x}{a} \exp(-\Gamma z). \tag{16}$$

Since E_y is the only component of electric field, the lines of force are normal to the guide faces at $y = 0, b$, and since E_y has been made to vanish when $x = 0, a$, there is no component tangential to the faces $x = 0, a$. It can be seen from the expressions for the field components that H_y is identically zero and that H_x vanishes on the walls $x = 0, a$. The boundary conditions are therefore satisfied.

Propagation Constant and Critical Wavelength

If the medium is loss-free, i.e. $g = 0$, equations (8) and (11) give for the propagation constant

$$\dot{\Gamma} = \sqrt{\left(\frac{n\pi}{a}\right)^2 - \left(\frac{2\pi}{\lambda}\right)^2} \tag{17}$$

since

$$\omega(\epsilon\mu)^{\frac{1}{2}} = 2\pi/\lambda, \tag{18}$$

where λ is the wavelength of a plane wave of angular frequency ω in the medium filling the waveguide.

If λ is small enough to make Γ imaginary then the field equations represent a pattern propagated in the z-direction but if Γ becomes real we have a stationary pattern which attenuates exponentially with distance. The critical wavelength which λ must not exceed if propagation is to take place is given by $\Gamma = 0$. Hence

$$\lambda_c = 2a/n. \tag{19}$$

In the case of propagation we have zero attenuation constant and $\Gamma = j\beta$ where

$$\beta = \frac{2\pi}{\lambda_c}\sqrt{\left(\frac{\lambda_c}{\lambda}\right)^2 - 1} \text{ radians per metre.} \tag{20}$$

In the case of attenuation the phase constant is zero and $\Gamma = \alpha$ where

$$\alpha = \frac{2\pi}{\lambda_c}\sqrt{1 - \left(\frac{\lambda_c}{\lambda}\right)^2} \text{ nepers per metre.} \tag{21}$$

The longest critical wavelength corresponds to $n = 1$ and is twice the length of the side perpendicular to the electric field. Fig. 69 shows the pattern which is obtained when the wavelength is shorter than critical and Fig. 70 shows the

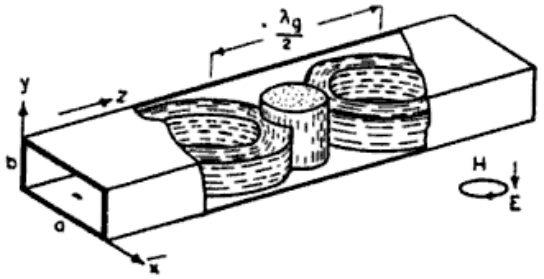


Figure 69.—Field pattern and coordinates for rectangular waveguide.

attenuated field for the same mode. The pattern of Fig. 69 is to be regarded as moving down the waveguide with a velocity $f\lambda_g$ which is somewhat above that of light. At fixed points in the median plane of the guide the electric and magnetic fields will be in phase, reaching their maxima together. At other points there is a phase difference which approaches 90 degrees at the walls $x = 0, a$. The patterns of Fig. 70

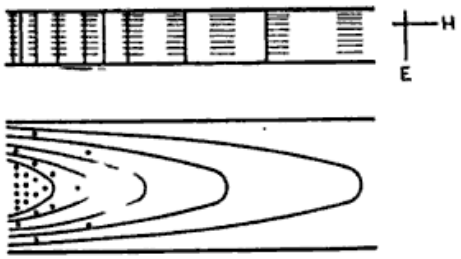


Figure 70.—Attenuated field of the usual mode in a waveguide too small to allow propagation.

are stationary in space but collapse and re-establish themselves harmonically with time. There is a 90-degree difference between the phases of the electric and magnetic fields so that when the electric field is as shown there is no magnetic field and vice versa. This field simply represents a periodic inter-

change of electric and magnetic energy without any net transfer. To amplify these remarks we note from (13) and (16) that $H_x = (-\Gamma/j\omega\mu)\, E_y$. Thus when Γ is imaginary the transverse field components are in phase and there is a net flow of energy but when Γ is real the fields are in quadrature and there is oscillation of stored energy only.

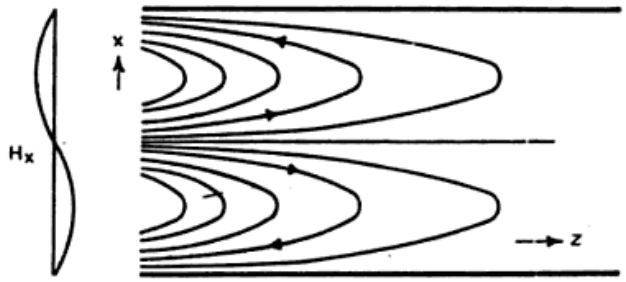


Figure 71.—Attenuated field of the TE_{10} mode in rectangular waveguide.

Fig. 71 indicates the meaning of the integer n which may be described as the number of half-period variations along the x-direction of any field component. The figure shows a field corresponding to $n=2$.

Phase Velocity

The propagation constant determines the wavelength in the guide,

$$\lambda_g = 2\pi j/\Gamma = \frac{\lambda}{[1-(\lambda/\lambda_c)^2]^{\frac{1}{2}}} \tag{22}$$

or in simpler form,

$$\frac{1}{\lambda^2} = \frac{1}{\lambda_c^2} + \frac{1}{\lambda_g^2}.$$

The phase velocity is given by

$$v_p = f\lambda_g = \frac{v}{[1-(\lambda/\lambda_c)^2]^{\frac{1}{2}}} = \frac{j2\pi v}{\Gamma\lambda} = \frac{j\omega}{\Gamma} \tag{23}$$

where $v = (\epsilon\mu)^{-\frac{1}{2}}$ is the velocity of a plane wave in the medium filling the guide.

It will be noticed that the wavelength in the guide exceeds the wavelength of a plane wave in the medium and that the phase velocity is greater than that of light.

Group Velocity

Propagation in a waveguide is dispersive, i.e. different frequencies travel at different velocities. The dispersion is normal since the velocity decreases with frequency. A group velocity may be calculated from the usual formula

$$u = v_p - \lambda \frac{dv_p}{d\lambda} \tag{24}$$

giving

$$u = \frac{v^2}{v_p}. \tag{25}$$

Since $v_p > v$ it follows that $u < v$ i.e. the group velocity is always less than the velocity of a plane wave in the medium.

From the field equations we can find the density of energy and flow of power and thus make a direct calculation of the velocity with which energy is transferred through the waveguide. The mean flow of power P down the guide is given by the real part of the flux of the complex Poynting vector $\mathbf{\Pi}$ through a transverse section.

$$P = Re \iint \mathbf{\Pi} \cdot \mathbf{d\Sigma} \tag{26}$$

where $d\Sigma$ is an element of area in a transverse section and

$$\mathbf{\Pi} = \tfrac{1}{2}\mathbf{E} \times \mathbf{H}^*. \tag{27}$$

Now

$$\begin{aligned} \Pi_z &= -\tfrac{1}{2}E_y H_x^* \\ &= \tfrac{1}{2}\frac{\Gamma E^2}{j\omega\mu}\sin^2\frac{n\pi x}{a}. \end{aligned} \tag{28}$$

Hence

$$\begin{aligned} P &= \int_0^a\int_0^b \Pi_z \, dxdy \\ &= \frac{(Eb)^2}{4j\omega\mu b/a\,\Gamma}. \end{aligned} \tag{29}$$

At a point where the electric field strength is E the density of electric energy is $\epsilon E^2/2$. As the electric field in the guide is distributed sinusoidally in two dimensions (the x- and z-directions) with a maximum strength E, the mean electric energy density taken over a section of guide $\lambda_g/2$ long is one quarter this value. Taking account of the equal quantity of magnetic energy we have for the total average energy density,

$$U = \frac{\epsilon E^2}{4}. \tag{30}$$

The energy transport velocity u can be found from

$$u = \frac{P}{abU}, \tag{31}$$

whence

$$u = \frac{\Gamma}{j\omega\mu\epsilon} = \frac{2\pi}{\omega\lambda_g\mu\epsilon} = \frac{v^2}{v_p}. \tag{32}$$

The velocity of energy flow is therefore less than the velocity of the wave and is equal to the group velocity.

Attenuation

To find the attenuation due to the finite conductivity of the walls we calculate the power flow into the walls. The attenuation in nepers per unit length is then one half the fraction of the transmitted power which is lost in unit length since $dP/dz = -2\alpha P$. From equation (13) which introduces the definition of attenuation constant we see that the neper, unlike the decibel, measures the attenuation of field or voltage but not power. When the two units are mutually convertible, one neper equals 8·686 decibels.

Assuming that the loss of energy does not appreciably modify the field distribution, the complex power flow into unit length of the walls is

$$\tfrac{1}{2}\int\int E_t H_t^* \, dS \tag{33}$$

where E_t and H_t are the tangential components of the field

and dS is an element of area of the walls. The small tangential electric field E_t is given by

$$E_t = \eta H_t \tag{34}$$

where η is the intrinsic impedance of the conductor. Since $\eta = [j\omega\mu(g + j\omega\epsilon)]^{\frac{1}{2}}$ and therefore in a conductor $\eta = (j\omega\mu/g)^{\frac{1}{2}}$ we have

$$E_t = \mathsf{R}(1 + j)H_t \tag{35}$$

where R the intrinsic resistance of the conductor is given by

$$\mathsf{R} = \left(\frac{\omega\mu}{2g}\right)^{\frac{1}{2}} \text{ ohms.} \tag{36}$$

The intrinsic resistance of a conductor is equal to the resistance to direct current of a square piece whose thickness is equal to the skin depth. Intrinsic resistance increases as the square root of the frequency and is least for good conductors.

We may obtain the skin depth δ from the relation

$$\mathsf{R} = \frac{1}{g\delta} \tag{37}$$

giving

$$\delta = \sqrt{\frac{2}{\omega\mu g}}. \tag{38}$$

Skin depth is defined so that the total current in the surface of a conductor, uniformly distributed to a depth δ, would dissipate the same power as the actual exponentially distributed current. Curves showing the skin depth for various conductor materials and frequencies are given in Fig. 106, Chapter VII. The skin depth in a conductor increases as the square root of the wavelength and is least for good conductors.

TABLE 4—CONDUCTIVITIES RELATIVE TO COPPER

Conductor	*Ag*	*Cu*	*Au*	*Al*	*Zn*	*Brass*
g_r	1·06	1·00	0·70	0·61	0·28	0·24
$g_r^{-\frac{1}{2}}$	0·97	1·00	1·19	1·28	1·90	2·04

For copper we have $\mathsf{R}_0 = 0{\cdot}00452\lambda^{-\frac{1}{2}}$ ohms and $\delta_0 = 3{\cdot}8\lambda^{\frac{1}{2}} \times 10^{-6}$ metres and for non-magnetic materials whose conductivity relative to that of copper is g_r we have $\mathsf{R} = \mathsf{R}_0 g_r^{-1}$ and $\delta = \delta_0 g_r^{-\frac{1}{2}}$. Table 4 gives the factor g_r for a few metals.

At a wavelength of 10 centimetres $\mathsf{R}_0 = 0{\cdot}0143$ ohms and $\delta_0 = 1{\cdot}2 \times 10^{-6}$ metres which is no more than twice the wavelength of light.

The average flow of power W into unit length of the walls is the real part of equation (33), hence

$$W = \frac{\mathsf{R}}{2} \iint H_t H_t^* \, dS. \tag{39}$$

The energy lost per cycle, $2\pi W/\omega$, is

$$\iint \pi\delta \, (\tfrac{1}{2}\mu H_t H_t^*) \, dS. \tag{40}$$

As $\frac{1}{2}\mu H_t H_t^*$ is the density of magnetic energy at the surface of the conductor this result means that in each cycle an amount of energy is dissipated equal to the magnetic energy stored in a skin of thickness $\pi\delta$. An idea of the order of magnitude of the attenuation in a waveguide may therefore be formed by comparing the area of this thin skin ($\pi\delta \times$ perimeter) with the cross-sectional area of the guide leading to

$$\alpha \doteqdot \frac{\pi\delta \times \text{perimeter}}{2 \times \text{area}} \left(\frac{\lambda_g}{\lambda}\right)^2 \text{nepers per guide-wavelength} \tag{41}$$

or, approximating further

$$\alpha \doteqdot \frac{2\delta \times \text{perimeter}}{\lambda \times \text{area}} \text{ nepers per metre.} \tag{42}$$

These formulae show clearly how waveguide attenuation depends on wavelength, conductor material and size and shape of guide and readily yield its order of magnitude.

Now splitting W into two integrals over the walls parallel and perpendicular to the electric field

$$\begin{aligned} W &= \mathsf{R} \int_0^b H_t H^* \, dy + \mathsf{R} \int_0^a H_t H_t^* \, dx \\ &= \frac{-\mathsf{R}\pi^2 b E^2}{\eta^2 I^2 a^2} + \frac{\mathsf{R} a E^2}{2\eta^2} \\ &= \frac{\mathsf{R} a E^2}{2\eta^2}\left[1 + \frac{2b}{a}\left(\frac{\lambda}{\lambda_c}\right)^2\right] \end{aligned} \tag{43}$$

where the intrinsic impedance of the medium in the guide is

$(\mu/\epsilon)^{\frac{1}{2}}$. The total power flowing through the guide is

$$P = \frac{ab\,\Gamma E^2}{4j\omega\mu}. \tag{44}$$

Hence the attenuation is

$$\alpha = \tfrac{1}{2}\frac{W}{P} = \frac{\pi\delta\left[\frac{a}{b} + 2\left(\frac{\lambda}{\lambda_c}\right)^2\right]}{a\lambda\left[1 - \left(\frac{\lambda}{\lambda_c}\right)^2\right]^{\frac{1}{2}}}. \tag{45}$$

Thus for a copper waveguide 3 inches by 1 inch internal dimensions this equation gives an attenuation of 0·00253 nepers per metre at a wavelength of 10 centimetres. From the approximate formula (41) we have 0·0026 nepers per metre.

Review

In the preceding paragraphs the properties of a simple waveguide have been worked out. Starting from Maxwell's equations and applying the boundary conditions for a rectangular conducting pipe we have derived a set of solutions [equations (13), (15), (16)] in each of which the electric field is everywhere parallel to the y-direction. Each of these modes has its critical wavelength given by equation (19). At shorter wavelengths than the critical the wave travels with a velocity given by equation (23) and at longer wavelengths equation (21) gives the attenuation constant governing its decay. The group velocity has been calculated from the known frequency-dependent wave velocity [equation (25)] and shown to be equal to the speed at which energy travels in the guide. This speed is always less than that of a plane wave in the medium whereas the wave itself always travels faster than a plane wave. By taking into account the energy spent in heating the walls of the waveguide, the attenuation constant has been found [equation (45)] and a useful approximate result given [equation (41)]. The latter shows that waveguide attenuation per wavelength is of the order of the ratio

of skin depth in the conductor to the linear dimension of the cross section. For a given waveguide the attenuation constant increases as the square root of the frequency but if the size of the waveguide is changed in proportion to the wavelength the attenuation constant increases as the sesquialterate power of the frequency. Field equations for the circular waveguide are given in section 7.

4. Synthesis of Waveguide Fields

The procedure just described for investigating the properties of a waveguide exemplifies the method of attack on waveguides of other shapes, i.e. Maxwell's equations are written out in a suitable coordinate system and solved for the appropriate boundary conditions. In the case of the rectangular waveguide carrying modes such as those we have been discussing, an alternative method is available which leads to a useful physical picture of the guided wave.

Consider an infinite conducting plane in a loss-free dielectric, upon which is incident an infinite plane wave whose electric vector is parallel to the plane (Fig. 72).

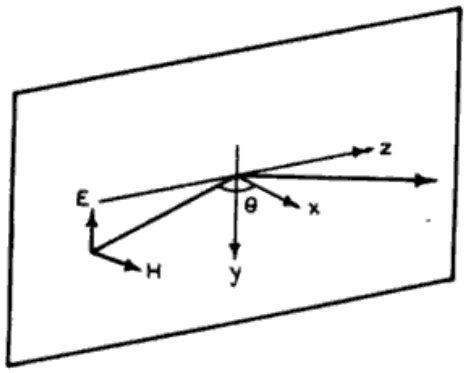


Figure 72.—Plane wave incident on a conducting sheet.

At the plane the electric field is annulled and the reflected wave sets up an interference pattern which moves without change along the plane.

Fig. 73 shows the crests and troughs of the electric field. There is no net flow of energy in the direction normal to the plane, but there is a steady flow parallel to the plane. If x is measured normal to the plane, y parallel to the electric vector and z is the direction in which the energy moves, the only field components are E_y, H_x and H_z. If λ is the wavelength of the incident radiation and θ is the angle of incidence, E_y will

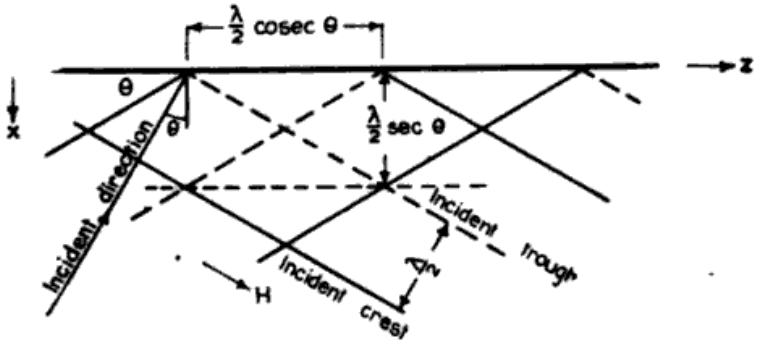


Figure 73.—Interference pattern at a conducting sheet. The electric vector is normal to and the magnetic field parallel to the plane of the diagram.

be zero on the planes $x = 0, \frac{1}{2}n\lambda \sec\theta$ parallel to the reflecting sheet. Consequently a conducting plane can be introduced into the field at a distance $\frac{1}{2}\lambda \sec\theta$ without disturbing the distribution. Moreover if conducting planes are introduced at $y = 0,b$ the boundary conditions will still be satisfied by the existing field.

We are now left with a travelling wave in a rectangular pipe. The field is the sum of two plane waves whose wavefronts make an angle θ with the axis, and which are reflected back and forth down the pipe. The distance measured along the pipe between points in the same phase exceeds the wavelength in the medium and from Fig. 73 is seen to be $\lambda \operatorname{cosec}\theta$, hence the phase velocity $v_p = v \operatorname{cosec}\theta$. If the width of the pipe $a = \frac{1}{2}\lambda \sec\theta$ is kept fixed, then as the wavelength of the radiation is increased the angle θ approaches zero, the phase wavelength approaches infinity and the direction of the component waves approaches normal incidence on the walls. When θ is zero, $\lambda = 2a$; consequently $2a$ is the critical wavelength which λ may not exceed. For very short wavelengths

θ approaches 90 degrees, the phase wavelength approaches the wavelength of a plane wave in the medium, and the direction of the component waves lies mainly along the axis. The energy velocity, as we have seen, is $v^2/v_p = v \sin \theta$ and this is the resolved part of velocity of the plane wave components along the axis. If the amplitude of the plane waves is $\frac{1}{2}E$ then the maximum electric field is E. The power carried by such a plane wave is $E^2/8\eta$ per unit area. If we double this, resolve along the axis and multiply by the area of the pipe, we have for the flow of power, $P = E^2 ab \sin \theta/4\eta = E^2 ab\lambda/4\eta\lambda_g$ which agrees with (44). We also observe that since $\lambda/\lambda_g = \sin \theta$ and $\lambda/\lambda_c = \cos \theta$ it follows that $(\lambda/\lambda_g)^2 + (\lambda/\lambda_c)^2 = 1$. The transverse components of electric and magnetic fields are in a constant ratio independently of the coordinates. For a plane wave $E/H = \eta$; in the waveguide, $-E_y/H_x = \eta/\sin \theta = \eta\lambda_g/\lambda = j\omega\mu/\Gamma$.

Most of the results for the rectangular waveguide have now been obtained by the method of decomposition into plane waves.

5. Nomenclature of Wave Types

To illustrate the rectangular waveguide modes other than those discussed here in detail, Figs. 74, 75 and 76 have been prepared.

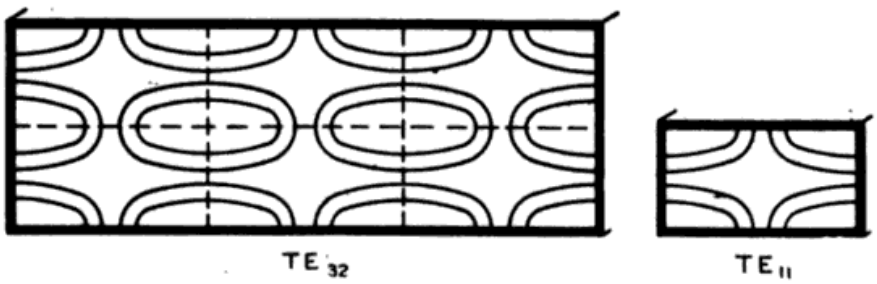


Figure 74.—Electric lines in transverse section of a waveguide carrying the TE_{32} mode. Conducting partitions may be inserted along the dashed lines without disturbing the field. Each cell then carries a TE_{11} wave.

The TE_{10} mode is the one with which we have been mainly concerned and the subscripts indicate that there is one half-period variation in the x-direction and none in the y-direction. The electric field in the transverse plane of the TE_{32} mode

(Fig. 74) shows three half-period variations in one direction and two in the other and we see that the wave may be regarded as divided into six equal cells in each of which a TE_{11} mode is present. The dividing walls between the cells occur at places where conducting planes could be inserted. To

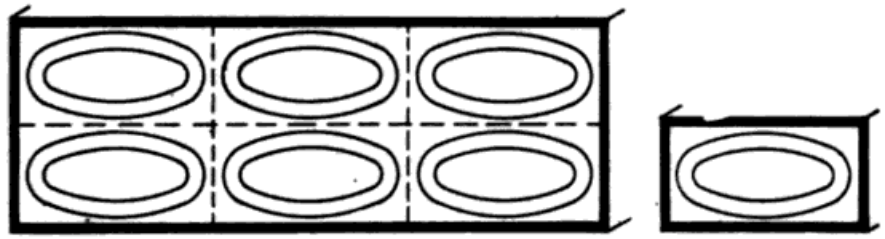


Figure 75.—Magnetic lines of the TM_{32} mode in rectangular waveguide. Each of the six cells carries a TM_{11} wave.

visualise the TM_{32} mode (Fig. 75) one should fill each cell with a TM_{11} wave. As neither of the subscripts of transverse magnetic waves can be zero, all modes in rectangular waveguides are built up from three simple modes only. These are the TE_{10}, TE_{11} and TM_{11}. A grasp of these three modes of propagation, together with the fundamental circular electric mode, will be found sufficient for forming a qualitative picture

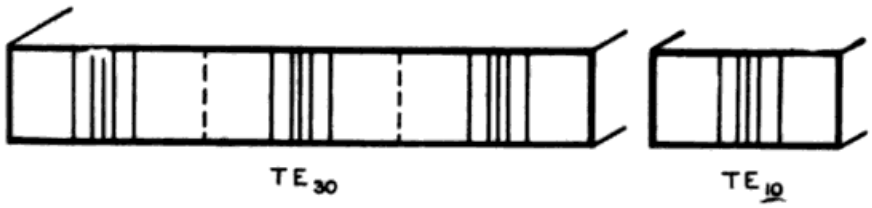


Figure 76.—Electric field of the TE_{30} mode. Each of the cells carries a TE_{10} wave.

of the modes of transmission in waveguides of circular or any other cross-section. To illustrate this Fig. 77 compares corresponding modes in rectangular and circular waveguides. The correspondence is such that one pattern may be transformed into the other by continuous deformation of the wall. The circular electric mode, which is also shown, has no analogue

in rectangular pipes—deformation of the pipe causes lines of force to snap and simple squeezing into a rectangular shape results in a TE_{20} mode.

The nomenclature of wave types in circular pipes is discussed in section 7 under circular cylindrical resonators.

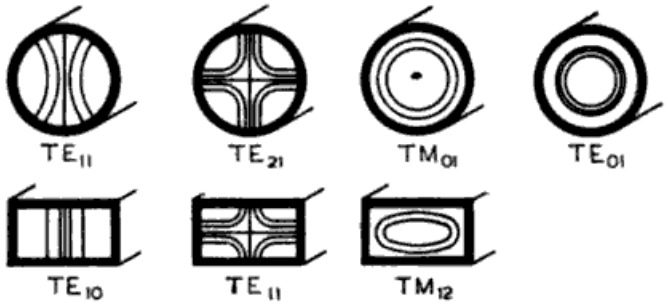


Figure 77.—Corresponding modes in circular and rectangular pipes.

6. Waveguide Impedance Theory

Normalised Impedances in Waveguides

Much of the foregoing matter has been concerned with emphasising the differences between waveguides and conventional transmission lines and these differences must be kept in mind. Yet there is a similarity between the two which is most important in practice for it allows the impedance-measurement techniques in use with conventional lines to be applied to waveguides. If, in a length of waveguide, discontinuities are so remote that the dominant mode exists in virtual purity, then there will in general be unequal waves travelling in opposite directions and giving rise to a standing wave pattern such as might be found on a conventional transmission line. The pattern is sufficiently specified by two real numbers: the standing wave ratio[2] S which is the ratio of the minimum

[2] Standing wave ratio so defined lies between 0 and 1. The reciprocal is sometimes used in which case standing wave ratio ranges from 1 to ∞. Two advantages are possessed by the present definition: on the Kennelly charts the standing wave ratio of every circle centred on the real axis is readily found as the abscissa of its intersection with that axis, on the Smith chart equal increments in standing wave ratio produce roughly equal increments in the radii of the circles of constant standing wave ratio, thus facilitating interpolation.

to the maximum field and the space phase θ which is the distance in angular measure of a minimum from a fixed reference point. Fig. 78 shows the standing wave on an actual transmission line. The two quantities S and θ may now be used

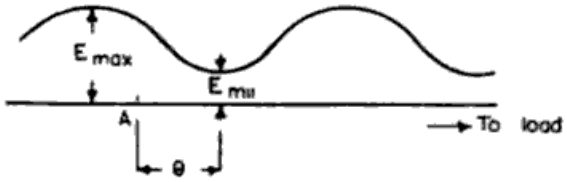


Figure 78.—Standing wave pattern on a transmission line.

to define the impedance at the point A by analogy with the conventional line for which

$$Z = Z_0 \phi\,(S, \theta) \tag{46}$$

where Z_0 is the characteristic impedance of the line. So far the question of the characteristic impedance of a waveguide has not arisen, nor need it here for it will be sufficient to have a definition of the normalised impedance ζ at a point in a waveguide

$$\zeta = \frac{Z}{Z_0} = \phi\,(S, \theta). \tag{47}$$

The actual form of the function ϕ is

$$\phi\,(S, \theta) = \tanh\,(\operatorname{artanh} S + j\theta). \tag{48}$$

It will be noticed that in this definition of impedance no ratio of voltages and currents is implied. We have simply defined a function of the observable quantities S and θ, which replaces the customary impedance of conventional line theory, when we are dealing with reflections and impedance matching.

Limitations of the Theory

As an example of the meaning of waveguide impedances, consider a metallic obstacle placed in a transverse plane of a waveguide. Suppose that a generator feeds one end of the waveguide which extends to infinity. Near the obstacle the field will be distorted in such a way that the electric lines terminate normally on the conducting surface. The total

field may be regarded as the sum of a number of the possible patterns which can exist in the undisturbed guide. Well away from the obstacle only the dominant mode will be found as no other mode is capable of propagation. Near the obstacle as many additional modes occur as are necessary to satisfy the boundary conditions and they are attenuated exponentially in both directions. The dominant mode will be found travelling in both directions between the generator and the obstacle but beyond the obstacle travelling outwards only. These are the incident, reflected and transmitted waves. As far as the behaviour of the dominant waves is concerned the system may be interpreted as a uniform line having a shunt reactance at the position of the obstacle. If the reactance is negative then the energy stored by the local fields will be predominantly electric.

The last paragraph indicates that the analogy with the conventional line has strict limitations. Two conditions which must be fulfilled are firstly that only one mode can be propagated, and secondly that attenuated modes shall be negligible, in the neighbourhood of the point where impedances are to be measured.

Impedance Charts

The bulk of numerical work with transmission lines is carried out graphically on charts of two kinds, the Kennelly chart, Fig. 79 and the Smith chart,[3] Fig. 80. Each of these consists of two sets of superimposed orthogonal networks, one of S and θ, the other of normalised resistance ($\rho = R/Z_0$) and reactance ($\xi = X/Z_0$). A point on the chart is therefore characterised by a pair of values of S and θ and the components of a complex impedance ζ which satisfies the relation $\zeta = \phi\ (S,\ \theta)$.

[3] A. E. Kennelly, *Chart Atlas of Complex Hyperbolic and Circular Functions*, Harvard University Press, Cambridge, Mass., 1924 (of this collection the hyperbolic tangent charts are the ones referred to here); and P. H. Smith, "Transmission line calculator," *Electronics*, Vol. 12, January, 1939, pp. 29-30.

To construct the Kennelly chart the relation

$$\zeta = \tanh \mu \tag{49}$$

where $\mu = \operatorname{artanh} S + j\theta$ (θ in quadrants), has been plotted on the complex plane of ζ.

By making the transformation

$$w = \frac{\zeta - 1}{\zeta + 1} \tag{50}$$

the Smith chart is obtained. In polar coordinates,

$$w = \frac{1 - S}{1 + S} \exp j2\theta.$$

As the Kennelly chart represents impedances in rectangular coordinates it is useful where lumped resistances and reactances occur and when calculations are carried out in terms of im-

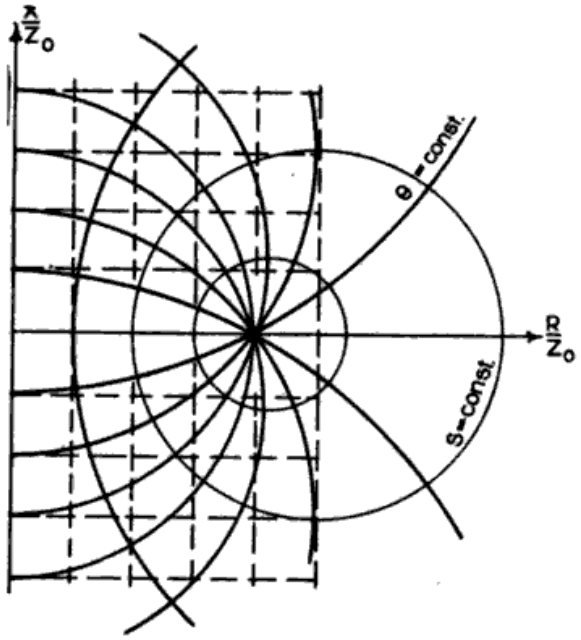


Figure 79.—Kennelly chart used for transmission line calculations.

pedances, as for example when changes in characteristic impedance have to be made. In some experimental work the measurable quantities S and θ may be sufficient for the calculations without at any stage changing to impedances.

Or again highly reactive impedances may be involved. In these cases the Smith chart has advantages for it is easier to plot on the S and θ loci which are radial lines on concentric

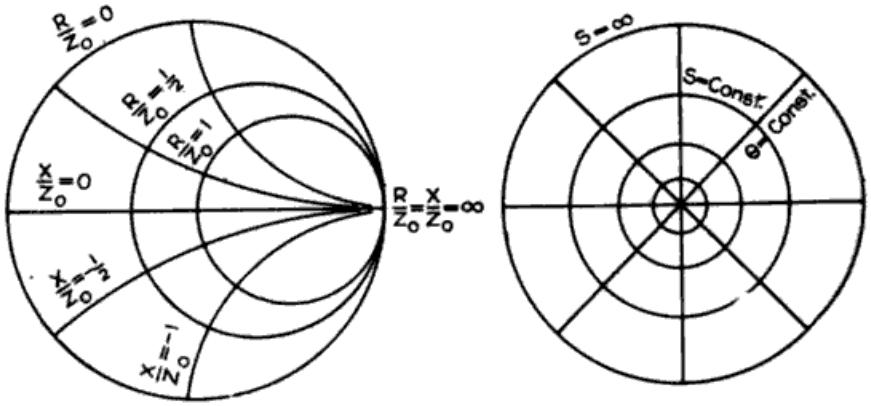


Figure 80.—Smith chart. *Left*—Resistance and reactance circles. *Right*—S and θ loci.

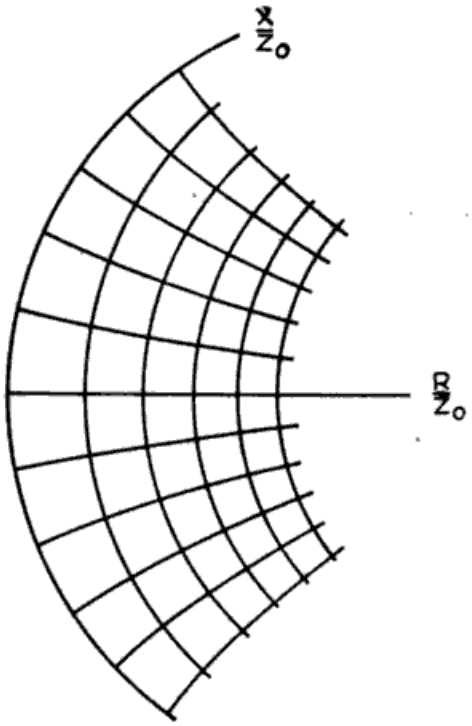


Figure 81.—Passing from the Kennelly to the Smith plane.

circles, and in the second case the chart covers the whole semi-infinite range of values of ζ. In general it is found that one chart usually offers advantages over the other when a particular calculation is to be done and it is necessary to be familiar with both. Fig. 81 suggests a way of picturing the

transition from the Kennelly to the Smith chart showing how the infinite half-plane is compressed into a circle.

The graphical charts are not only used for numerical calculations. A good deal of qualitative thinking can be done quickly and easily with their aid which could otherwise hardly be done at all. Examples of this use are given in the next chapter.

Wave Impedance and Power Flow

So far the measurements discussed have been independent of the absolute level of the electromagnetic fields involved. Sometimes however one wishes to calculate the electric or magnetic field when the flow of power is given or to calculate one field component in terms of the other. This requirement introduces the idea of wave impedance which in a waveguide is defined as the ratio of the transverse field components. If u and v are mutually perpendicular directions in the transverse plane then the wave impedance is given by

$$K_z = \frac{E_u}{H_v} = -\frac{E_v}{H_u}. \tag{51}$$

The senses of u and v are chosen to make K_z positive for a positive-going wave. In a rectangular waveguide carrying fields of the type discussed above, we have from equations (13) and (16)

$$K_z = -\frac{E_y}{H_x} = \frac{j\omega\mu}{\Gamma} = \frac{\eta}{\left[1 - \left(\frac{\lambda}{\lambda_c}\right)^2\right]^{\frac{1}{2}}}. \tag{52}$$

From equation (29) the flow of power P down a rectangular waveguide is

$$P = \frac{a}{4b}\frac{VV^*}{K_z} \tag{53}$$

where $V = Eb \exp(j\omega t - \Gamma z)$ is the integral of the electric field in the medium plane from one wall of the waveguide to the other. By comparing this equation with

$$P = \frac{VV^*}{2Z_0} \tag{54}$$

for a conventional transmission line one is tempted to define

the characteristic impedance of a waveguide by the formula

$$Z_0 = \frac{2b}{a} K_z. \tag{55}$$

Such a course cannot lead to error but if subsequently an attempt is made to find an analogue for I, the current in a conventional line, care will be needed. Suppose as is commonly done, that the total longitudinal current

$$\int_0^a H_x dx = -\frac{2\Gamma a}{j\omega\mu\pi} E \exp(j\omega t - \Gamma z)$$

crossing a transverse line in the wide wall is called I. Then the two relations between P and I and between V and I will define two more characteristic impedances equal neither to the one already obtained nor to one another. One possibility is to have three different impedances but as they differ only by a numerical factor the course suggested here is to retain one only, the first mentioned, and modify the usual equations containing I.

Thus

$$P = \frac{\pi^2}{16} \cdot \frac{Z_0 II^*}{2} \tag{56}$$

$$V = \frac{\pi}{4} Z_0 I. \tag{57}$$

It is reasonable to make the choice in favour of the equation containing P and V for the quantity V has important significance relating to dielectric failure whereas I might with equal or more utility have been defined in terms of the maximum longitudinal current density as $a(H_x)_{max}$. The advantage of the above procedure lies in having a definite though arbitrary characteristic impedance instead of three or none. It should be stressed that the characteristic impedance has been agreed upon only for the rectangular waveguide carrying the TE_{10} mode but that in general a unique wave impedance is always

available. The characteristic impedance has two uses: (*a*) for simply relating potential difference and power and (*b*) for calculating reflections at a junction between two TE_{10} waveguides of different dimensions.

A 3-inch × 1-inch waveguide has a wave impedance of 506 ohms and a characteristic impedance of 337 ohms at a wavelength of 10 centimetres. If one megawatt is flowing the maximum voltage is $2{\cdot}6 \times 10^4$ volts and the gradient $1{\cdot}02 \times 10^6$ volts per metre. The maximum longitudinal current is 98 amperes and the maximum current density $1{\cdot}69 \times 10^9$ amperes per square metre for a copper waveguide.

7. Cavity Resonators

Solution of Maxwell's Equations

A dielectric region completely enclosed by a conducting surface constitutes a cavity resonator. Such a resonator possesses an infinite number of natural frequencies and corresponding field patterns. We shall first consider the case of perfectly conducting walls and non-conducting dielectric and then take account of damping.

The differential equation which the electric field in a resonator (or waveguide) must satisfy is obtained from Maxwell's equations as follows:

$$\begin{aligned} \text{curl } \mathbf{E} &= -j\omega\mu\mathbf{H} \\ \text{curl } \mathbf{H} &= j\omega\epsilon\mathbf{E}. \end{aligned} \tag{58}$$

Taking the curl of the first and substituting the second,

$$\text{curl curl } \mathbf{E} = \omega^2\epsilon\mu\mathbf{E} \tag{59}$$

since curl curl = grad div $- \nabla^2$ and since div $\mathbf{E} = 0$ we have

$$\nabla^2\mathbf{E} + \omega^2\epsilon\mu\mathbf{E} = 0. \tag{60}$$

This is the vector wave equation. Solutions to this equation are obtained by separation of variables. In a limited number of coordinate systems (about five all told) the vector wave

equation splits into three total second-order differential equations and the field may be expressed as the product of the three functions satisfying these equations. Although direct solution by this method can be carried out for comparatively few shapes of cavity its value lies in the possibility of constructing further solutions by linear superposition. The problem is similar to that of acoustical resonance though more difficult.

Once a solution for **E** has been found **H** can be obtained from

$$\mathbf{H} = \frac{j \operatorname{curl} \mathbf{E}}{\omega \mu} \tag{61}$$

and will be found to satisfy the necessary boundary conditions, viz. that there should be no normal component of magnetic field and no tangential component of electric field on a perfectly conducting surface. For over a closed path in the surface, $\int \mathbf{E} \cdot \mathbf{ds} = 0$ since there is no tangential component of **E**. Therefore $\iint \operatorname{curl} \mathbf{E} \cdot \mathbf{ds} = 0$. Hence the normal component of curl **E** vanishes on the surface and **H** is everywhere tangential.

Rectangular Prism

Consider a rectangular prism whose faces are at $x = 0,a$ $y = 0,b$, $z = 0,c$. By writing the vector wave equation in Cartesian coordinates, separating variables and determining the constants of integration from the boundary conditions viz. $E_x = 0$ when $y = 0,a$ and $z = 0,c$, etc., we have for the various components

$$E_x = -\frac{k_1 k_3}{k_1^2 + k_2^2} E \cos k_1 x \sin k_2 y \sin k_3 z$$

$$E_y = -\frac{k_2 k_3}{k_1^2 + k_2^2} E \sin k_1 x \cos k_2 y \sin k_3 z$$

$$E_z = E \sin k_1 x \sin k_2 y \cos k_3 z \tag{62}$$

$$H_x = j\omega\epsilon \frac{k_2}{k_1^2 + k_2^2} E \sin k_1 x \cos k_2 y \cos k_3 z$$

$$H_y = -j\omega\epsilon \frac{k_1}{k_1^2 + k_2^2} E \cos k_1 x \sin k_2 y \cos k_3 z$$

$$H_z = 0.$$

where

$$k_1 = l\pi/a, \quad k_2 = m\pi/b, \quad k_3 = n\pi/c, \tag{63}$$

$$(2\pi/\lambda)^2 = k_1^2 + k_2^2 + k_3^2, \tag{64}$$

and l, m and n are integers.

It will be noticed that H_z is zero. One further set of linearly independent solutions exists for which $E_z = 0$. The two types of solution are called Transverse Magnetic (TM) and Transverse Electric (TE) modes. The three integers (l, m, n) generate a triple infinity of each kind of mode. The resonant frequency of each mode is determined from the formula

$$\lambda = \frac{2}{\left[\left(\frac{l}{a}\right)^2 + \left(\frac{m}{b}\right)^2 + \left(\frac{n}{c}\right)^2\right]^{\frac{1}{2}}} \tag{65}$$

for the modes TE_{lmn} and TM_{lmn}. The longest wavelength is obtained when one of the integers is zero and the two associated with the largest dimensions are unity. In this case the electric field is parallel to the least dimension which then plays no part in determining the resonant frequency, i.e. the field is two-dimensional. Only in rectangular prisms is there complete degeneracy between TE and TM modes. In practice slight departures of the shape from squareness will cause the modes to separate. Another type of degeneracy occurs if two dimensions are equal e.g. if $a=b$. Then the modes TE_{pqr} will have the same frequencies as TE_{qpr}. Further degeneracies occurring for particular shapes are called accidental. As smaller and smaller wavelengths are considered the number of resonances which can occur in a resonator becomes very large. For a rectangular prism the number of modes resona-

ting at a wavelength greater tham λ_0 is about $8 \times \text{volume}/\lambda_0^3$. Thus a three-foot cube contains about 6000 modes with wavelengths greater than 10 centimetres. A similar qualitative

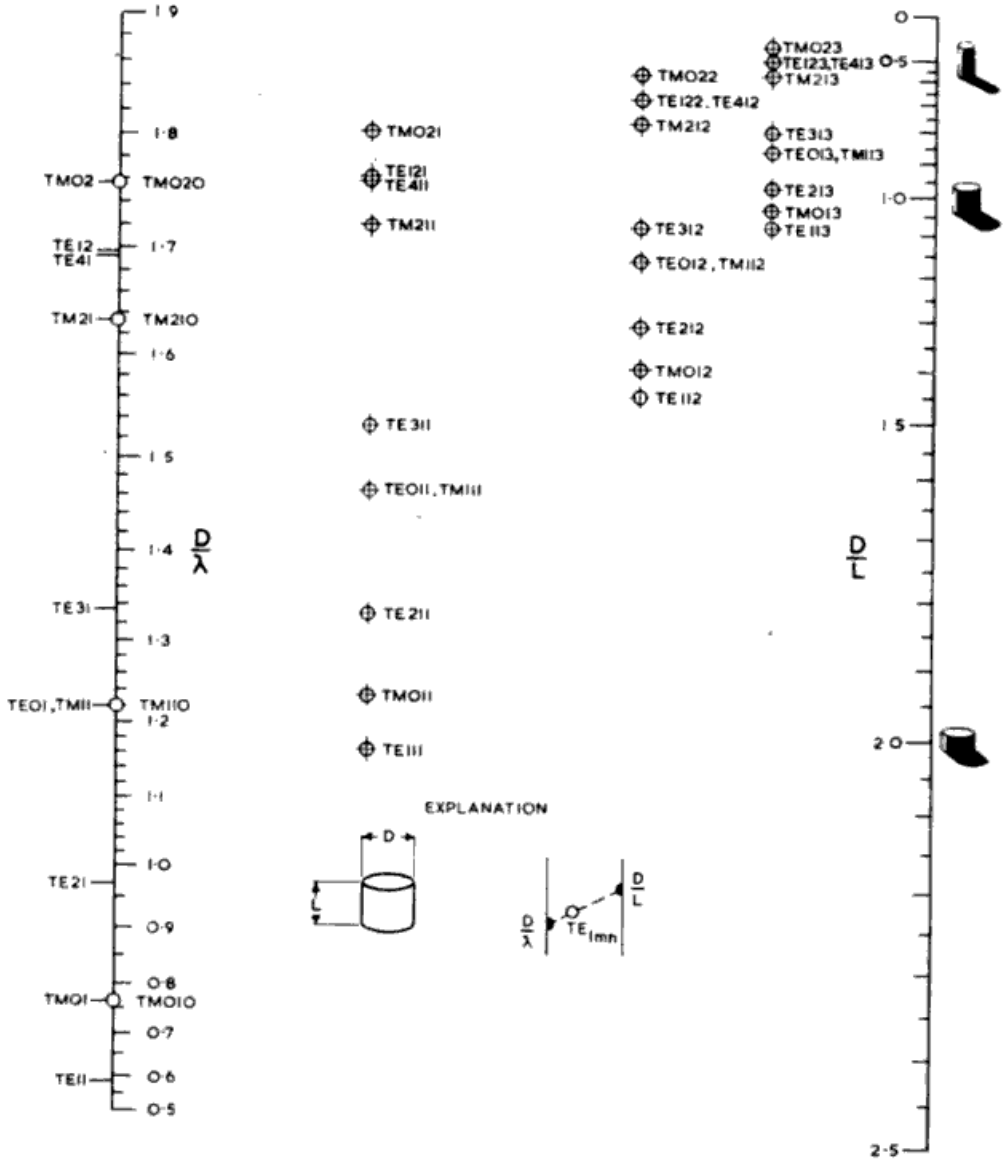


Figure 82.—Mode lattice for cylinder resonators.

situation holds for resonators of any shape. The distribution of modes in a circular cylinder can be obtained from Fig. 82 which is a mode lattice[4] suitable for numerical calculations.

[4] R. N. Bracewell, "Charts for resonant frequencies of cavities," *Proc. I.R.E.*, Vol. 35, 1947, pp. 830-41.

The field components of transverse electric waves in a rectangular prism are :

$$\begin{aligned} H_x &= -\frac{k_1 k_3}{k_1^2 + k_2^2} H \sin k_1 x \cos k_2 y \cos k_3 z \\ H_y &= -\frac{k_2 k_3}{k_1^2 + k_2^2} H \cos k_1 x \sin k_2 y \cos k_3 z \\ H_z &= H \cos k_1 x \cos k_2 y \sin k_3 z \\ E_x &= j\omega\mu \frac{k_2}{k_1^2 + k_2^2} H \cos k_1 x \sin k_2 y \sin k_3 z \\ E_y &= -j\omega\mu \frac{k_1}{k_1^2 + k_2^2} H \sin k_1 x \cos k_2 y \sin k_3 z \\ E_z &= 0. \end{aligned} \tag{66}$$

Some combinations of the integers, l, m, n, lead to trivial solutions or to fields incompatible with the boundary conditions. There are no resonances unless the following conditions are fulfilled : $l + m > 0$ for TM modes and $(l + m)n > 0$ for TE modes.

Circular Cylinder

The field components for the circular cylinder are given for the TM_{lmn} mode by

$$\begin{aligned} E_r &= -\frac{n\pi}{kh} E J_l'(kr) \cos l\theta \sin \frac{n\pi z}{h} \\ E_\theta &= \frac{n\pi l}{k^2 r h} E J_l(kr) \sin l\theta \sin \frac{n\pi z}{h} \\ E_z &= E J_l(kr) \cos l\theta \cos \frac{n\pi z}{h} \\ H_r &= -\frac{j\omega\epsilon l}{k^2 r} E J_l(kr) \sin l\theta \cos \frac{n\pi z}{h} \\ H_\theta &= -\frac{j\omega\epsilon}{k} E J_l'(kr) \cos l\theta \cos \frac{n\pi z}{h} \\ H_z &= 0. \end{aligned} \tag{67}$$

where (r, θ, z) are cylindrical polar coordinates, l, m and n

are integers, h is the height and a the radius of the cylinder and $x_{lm} = ka$ is the mth root of $J_l(x) = 0$. The resonant wavelength is given by

$$\left(\frac{a}{\lambda}\right)^2 = \left(\frac{x_{lm}}{2\pi}\right)^2 + \frac{n^2}{4}\left(\frac{a}{h}\right)^2. \tag{68}$$

For the TE_{lmn} mode we have the following field components :

$$\begin{aligned}
H_r &= \frac{n\pi}{kh} HJ_l'(kr) \cos l\theta \cos \frac{n\pi z}{h} \\
H_\theta &= -\frac{n\pi l}{k^2 rh} HJ_l(kr) \sin l\theta \cos \frac{n\pi z}{h} \\
H_z &= HJ_l(kr) \cos l\theta \sin \frac{n\pi z}{h} \\
E_r &= \frac{j\omega\mu l}{k^2 r} HJ_l(kr) \sin l\theta \sin \frac{n\pi z}{h} \\
E_\theta &= \frac{j\omega\mu}{k} HJ_l'(kr) \cos l\theta \sin \frac{n\pi z}{h} \\
E_z &= 0.
\end{aligned} \tag{69}$$

In this case $x_{lm} = ka$ is the mth root of $J_l'(x) = 0$. Table 5 gives a few values of these roots.

TABLE 5—ROOTS OF BESSEL FUNCTIONS

Mode	x_{lm}
TE_{11}	1·8412
TE_{21}	3·0543
TE_{01}	3·8317
TE_{31}	4·2014
TE_{41}	5·3175
TE_{12}	5·3315
TM_{01}	2·4048
TM_{11}	3·8317
TM_{21}	5·1356
TM_{02}	5·5201

The integers l, m, n, specifying the mode may be described as follows :

l = the number of full period variations undergone by any non-zero field component as θ varies from 0 to 2π.

m = the number of zeros of angular component of electric field in the case of TE waves or of axial component of electric

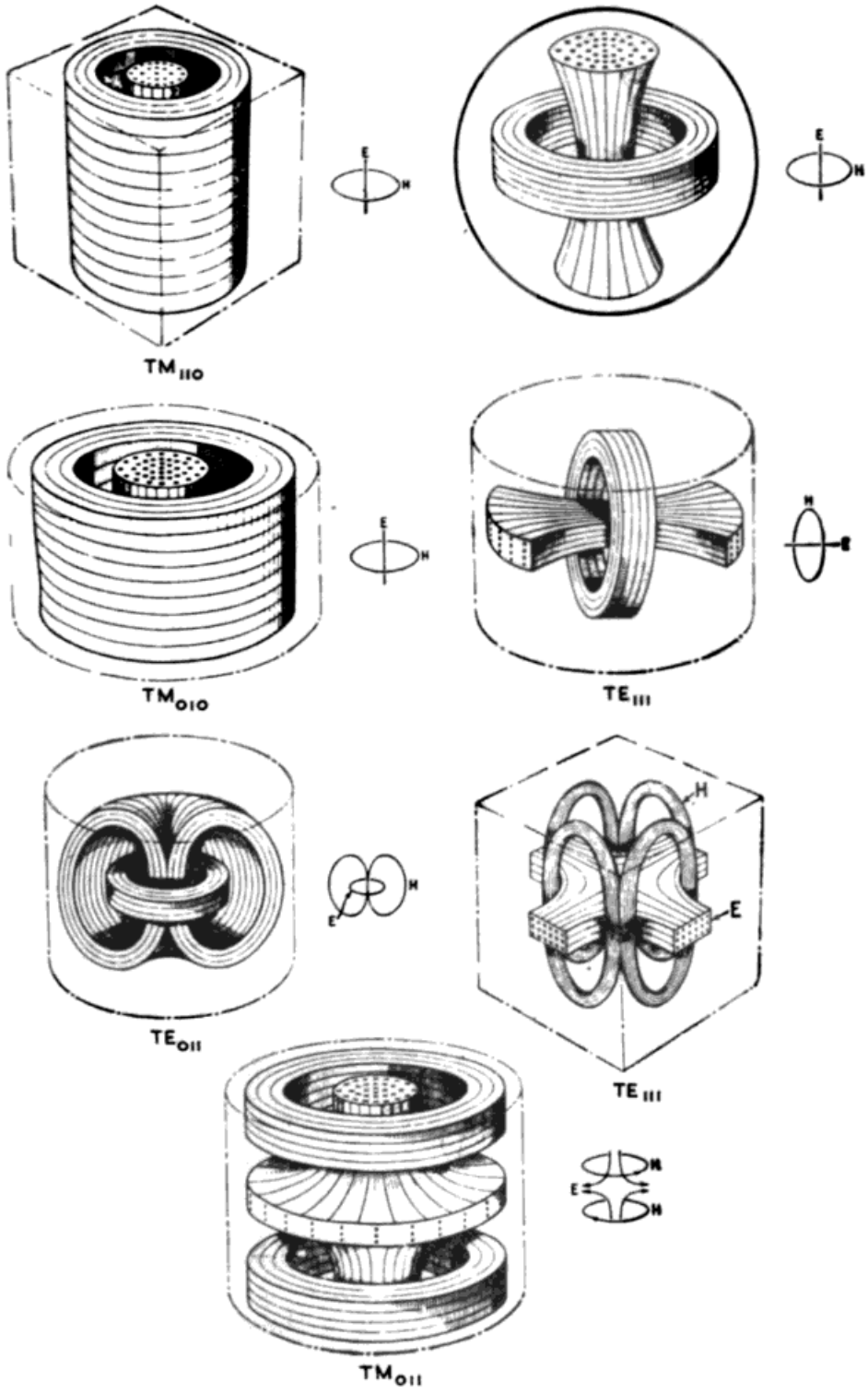


Figure 83.—Field patterns in cavity resonators.

field in the case of *TM* waves, lying on a radius, counting that at the wall but not that which may occur at the centre.

n = the number of half-wave variations through which the standing wave pattern extends in the z-direction.

Not all integral values may be assumed by l, m and n. Thus m cannot be zero. In the case of *TE* modes n cannot be zero, for this would involve a magnetic field parallel to the axis, which is not consistent with the boundary condition for magnetic fields on the conducting discs closing the cavity.

To calculate resonant frequencies for the circular cylinder the quickest procedure is to use the mode lattice shown in Fig. 82.

Field patterns of simple modes in various cavities are illustrated in Fig. 83. The drawings give three-dimensional views of the electric and magnetic fields which bring out some general properties of standing wave patterns in resonators, viz. that where the magnetic field is strong, the electric field is weak and that the magnetic lines form loops enclosing the electric lines. It can be seen that the magnetic lines run tangentially to the surface whilst the electric lines terminate normally. In one case, the TE_{011} mode in the circular cylinder, electric lines are shown which do not terminate at all but are closed on themselves. The phase relation of the two fields is such

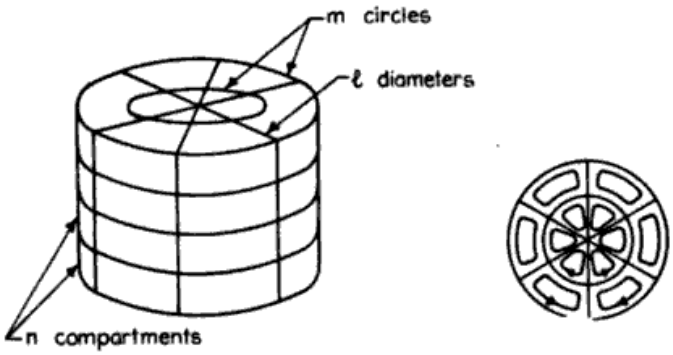


Figure 84.—The subscript notation for circular cylinders. The TM_{321} mode is used as a model and H-lines in the transverse plane are shown.

that the stored energy passes from one to the other each quarter-cycle or one is zero when the other is at a maximum.

It is of help in picturing a given mode in a circular cylinder to know that metallic partitions can be inserted in the resonator

without disturbing the field and that the cells into which the resonator is thus divided contain fields resembling the TE_{011}, TE_{111}, TM_{010} or TM_{011} modes all of which are shown in Fig. 83. The partitions are planes containing the axis, planes normal to the axis and coaxial cylinders. Altogether there are l nodal diametral planes, m nodal circles and n axial com-

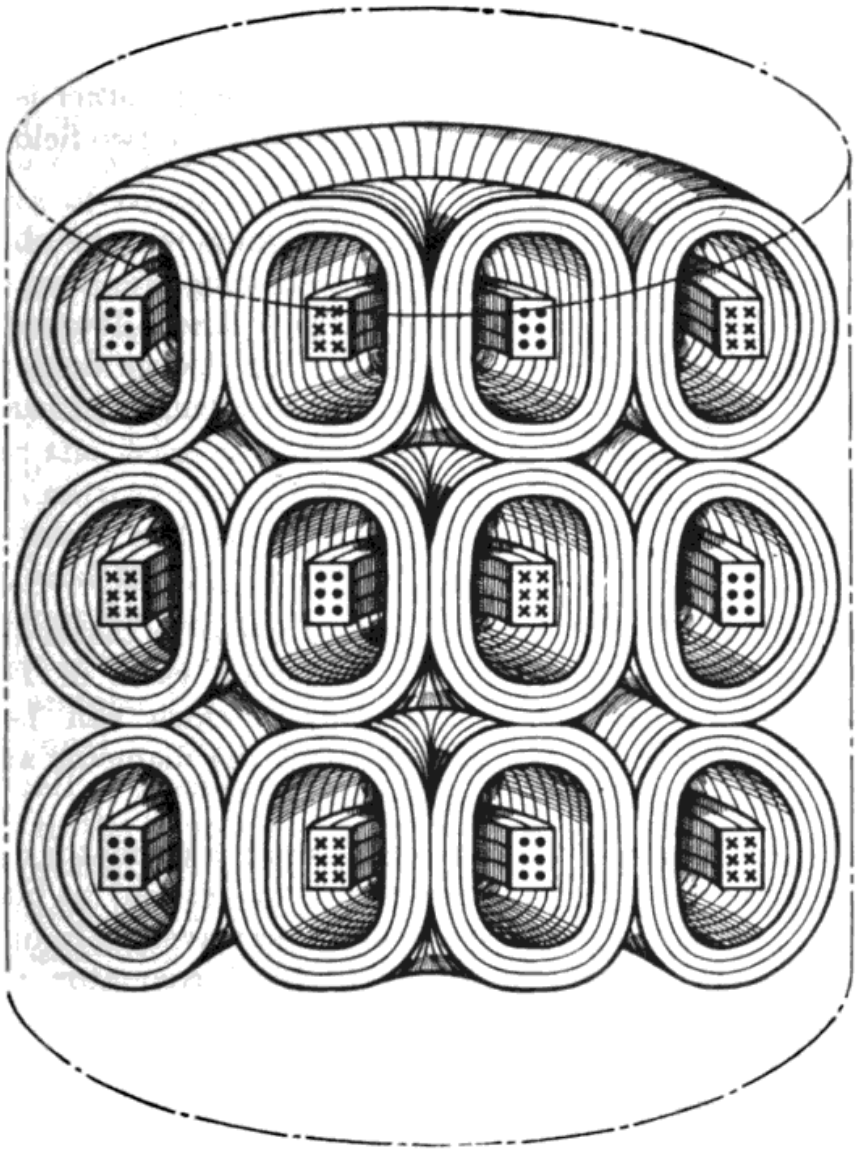


Figure 85.—The TE_{023} mode in a cylindrical cavity.

partments. This is illustrated in Fig. 84 which shows the partitions for the TM_{324} mode and the H-lines in the transverse plane. Each cell contains a field resembling the TM_{011} mode. Fig. 85 shows a cylinder containing the TE_{023} mode.

Travelling waves in circular pipes are named just as in circular resonators save that only the first two subscripts are necessary. Travelling wave patterns, such as that of Fig. 69, may be obtained from resonator patterns by displacing the electric field one quarter-wavelength axially. Comparison of the TE_{10} travelling wave Fig. 69 and the TE_{101} standing wave in a rectangular pipe, Fig. 83, shows the space relation between electric and magnetic fields in the two cases. In the first, one field type is at a maximum where the other is zero ; in the second, the transverse components of the two field types are everywhere in the same proportion.

Field equations for the circular waveguide may be obtained from equations (67) and (69) by replacing cos $n\pi z/h$ and sin $n\pi z/h$ by $\exp(-j2\pi z/\lambda_g)$ and $j\exp(-j2\pi z/\lambda_g)$ respectively and multiplying by j the expressions for electric field, where $(\lambda/\lambda_g)^2 + (\lambda/\lambda_c)^2 = 1$ and the critical wavelength λ_c is related to the diameter of the pipe by

$$\frac{D}{\lambda_c} = \frac{x_{lm}}{\pi}. \tag{70}$$

Table 5 allows D/λ_c to be evaluated or it may be taken directly from the D/λ-scale of Fig. 82. The least value of D/λ_c is 0·586. Hence the longest wavelength which can be propagated in a circular pipe is 1·70 times the diameter and the mode is the TE_{11}.

Instead of deriving the standing waves in a resonator directly it is possible (in cavities of constant cross-section) to build up the fields by superposition of equal waves travelling in opposite directions. Thus the circular cylindrical resonator may be regarded as a circular waveguide closed by plane ends an integral number of guide-wavelengths apart and the field as due to successive reflections of a travelling wave.

Modes in resonators formed of coaxial cylinders closed by annular ends are named after the corresponding cylindrical modes into which the patterns pass continuously as the inner cylinder shrinks to zero radius. The *TEM* modes however, cannot exist in the absence of the inner cylinder.

Coupling[5]

The cavities so far considered have been completely enclosed in a perfectly conducting envelope. We have shown that if any electromagnetic energy resides within certain cylindrical and prismatic cavities then it must be stored in one or more of the normal modes at the discrete natural frequencies of the cavity. If the slight spread due to finite conductivity of the walls is neglected, there can be no fields within the cavity, at frequencies other than the natural frequencies, in the absence of a generator. To sustain the field in a resonator a source of power must be introduced. Usually one desires to retain the properties of an undisturbed cavity so a small orifice in the wall is made and a field maintained across it resembling the field which would be there in the presence of the mode which it is desired to excite. Let the impressed fields at the point of coupling be $\mathbf{E}_0 \exp j\omega t$ and $\mathbf{H}_0 \exp j\omega t$ and let the fields at this point due to unit energy stored in the nth natural mode be $\mathbf{E}_n \exp j\omega_n t$ and $\mathbf{H}_n \exp j\omega_n t$. Then every mode will be excited for which the scalar products $\mathbf{E}_0 \cdot \mathbf{E}_n$ and $\mathbf{H}_0 \cdot \mathbf{H}_n$ do not vanish. This does not mean that oscillations are set up at the various natural frequencies, but that the various natural patterns execute forced vibrations at the angular frequency ω. The amplitude of excitation is proportional to the scalar products above but if $\omega \neq \omega_n$, it is small. Often only the mode whose natural frequency is nearest to the impressed frequency need be considered. When the impressed frequency is very close to a natural frequency, the amplitude of excitation of that mode is enormous in comparison with all others and limited by the damping due to finite conductivity of the walls.

It is worth while stressing the difference between free and

[5] Coupling is discussed in the following: E. V. Condon, "Forced oscillations in cavity resonators," *J. Appl. Phys.*, Vol. 12, 1941, pp. 129-32; S. A. Schelkunoff, "Representation of impedance functions in terms of resonant frequencies," *Proc. I.R.E.*, Vol. 32, 1944, pp. 83-90; C. G. Montgomery, R. H. Dicke and E. M. Purcell (Eds.), *Principles of Microwave Circuits* (Massachusetts Institute of Technology, Radiation Laboratory Series, Vol. 8), McGraw-Hill, New York, 1948.

forced vibrations in a cavity. A cavity with no source of power can only contain the natural field patterns at their respective natural frequencies whereas a cavity continuously excited by a simple harmonic oscillator contains, in addition to possible natural oscillations, any natural field pattern varying at the oscillator frequency. Figs. 67 and 68 show fields of the TM_{01} mode in circular waveguides of different sizes. In one case the fields extend with unchanged amplitude along the guide and in the other they are attenuated rapidly with distance. Forced patterns in a resonator behave in either of these ways and may therefore differ from the corresponding natural pattern to the extent that dependence on one coordinate may resemble a damped wave rather than a standing wave with nodes and antinodes. Some will make their presence felt throughout the cavity whereas others will be localised in the vicinity of the coupling point.

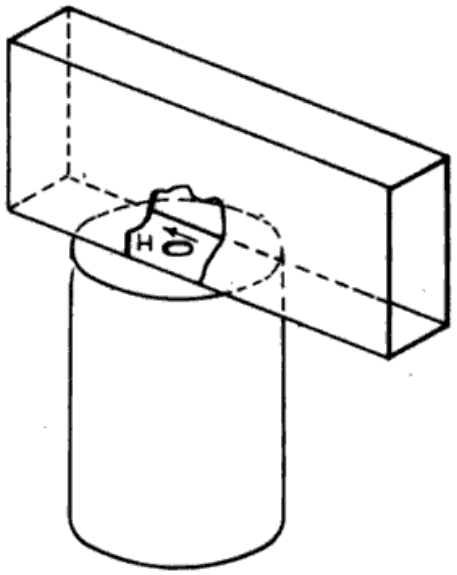


Figure 86.—Window-coupling to a cylindrical cavity.

Practical means of exciting cavities are divided into electric and magnetic according to the predominating component of the desired mode at the point of coupling. Both electric and magnetic coupling may be effected by feeding the cavity through a hole in a waveguide. Fig. 86 shows a cylindrical

cavity excited from a rectangular waveguide. At the coupling orifice there is normally no electric field so the coupling is magnetic. The magnetic field in the guide is tangential to the wall in the direction shown and any natural mode of the cavity, which calls for such a magnetic field at the orifice, will tend to be excited. A TE_{10} wave in the waveguide of Fig. 86 would be suitable for exciting the TE_{111} mode in the circular cylinder. Fig. 83 illustrates the latter mode.

Fig. 87 illustrates two methods which are commonly used

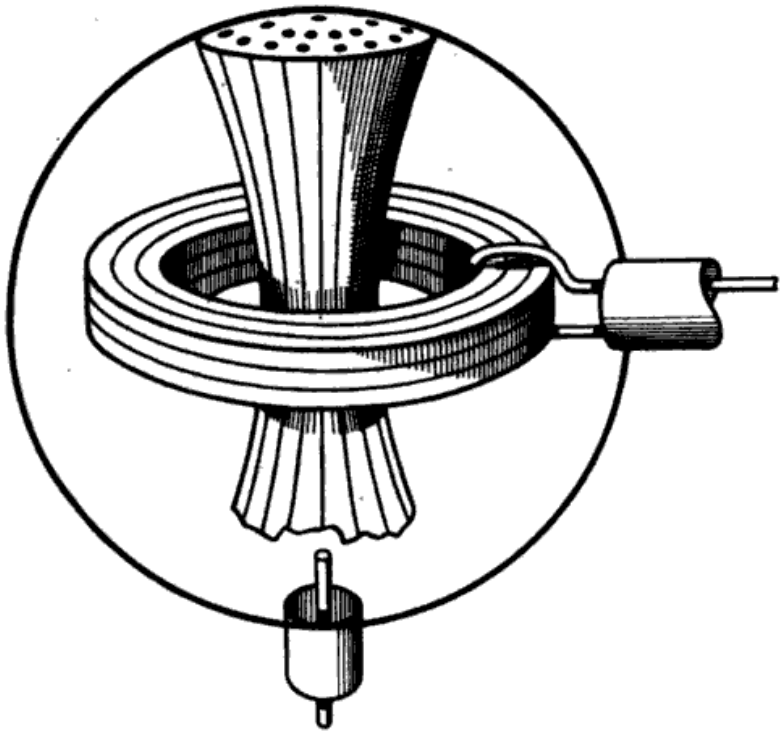


Figure 87.—Means of coupling to cavity resonators.

for coupling a coaxial transmission line to a cavity. The magnetic form of coupling comprises a loop which links with magnetic lines of force. The electric coupling may be regarded as a small aerial exciting modes which have a normal component of electric field. In the last case the degree of coupling may be controlled by withdrawing the probe and in the first, one may simply rotate the loop. If only one mode is excited two positions of zero coupling will exist, but unless the loop is situated at a point free from electric field, the

positions of zero coupling are not in general 180 degrees apart nor are the amplitudes of excitation equal at the two positions of maximum coupling.

Tuning

Adjustments to the resonant frequency of a cavity resonator are conveniently made by making changes in the shape. A common way of doing this is to insert plugs through the walls. Often the effect on resonant frequency cannot be simply calculated but it is possible to say whether the frequency will be increased or decreased from a knowledge of the undisturbed fields. If a conducting object is introduced into the resonator at a place where the electric field predominates then the resonant frequency will fall (Fig. 88). The frequency

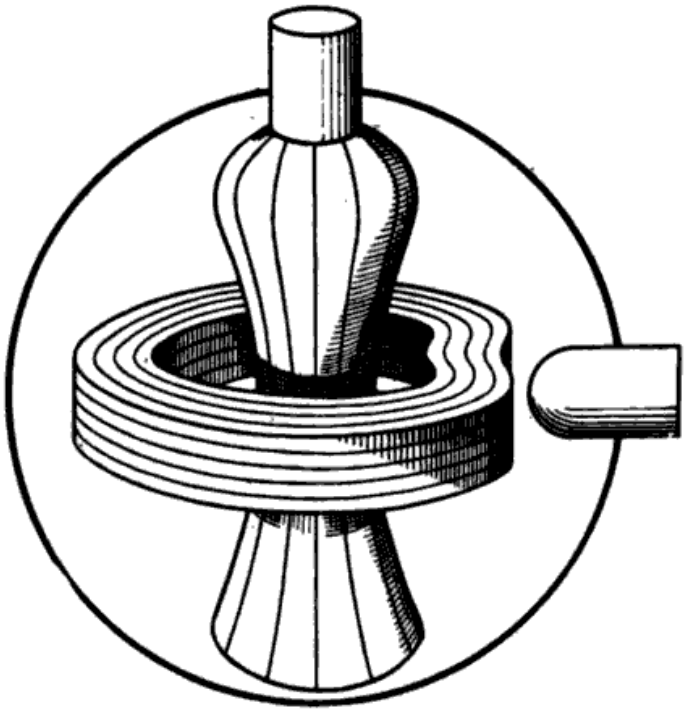


Figure 88.—Effect of tuning slugs on resonant frequency.

will rise if the object is placed in a strong magnetic field. A resonant coil and condenser show the same effect, for if metal is placed between the condenser plates the capacity increases but metal in the coil lowers its inductance. To estimate the

order of magnitude of the change we may assume that the proportional change in resonant frequency is equal to the average proportion of the total energy of the resonator which is stored in the electric or the magnetic form within the volume to be occupied by the plug. If the plug is to be inserted where there is both electric and magnetic energy, then the difference between the two average values must be taken. It follows that there are places in the walls of most resonators where plugs may be inserted without much change of frequency.

For small plugs and small insertions the effect is roughly proportional to the volume of the plug and if the point is one of pure magnetic or electric field the proportional change in wavelength is approximately equal to the ratio of the volume of the plug to the volume of the cavity. A perturbation theory for obstacles in cavities is given by Slater in the paper mentioned in the previous section. The approximate results given above represent the first term of a series developed by Slater.

Damping

The damping of fields in resonators results from energy loss due to the conductivity of the walls and dielectric. Where low damping is required dielectric loss can be rendered negligible in comparison with conductor loss by using air dielectric. Where both sorts of loss are present the resultant damping is the sum of the separate dampings. If the conductivity of a uniform dielectric completely filling the resonator is g and its permittivity ϵ, then the damping is simply $g/\omega\epsilon$ for any shape or mode. In the more important case of conductor loss however the shape and mode are very important. It is usual to work in terms of the reciprocal of the damping viz. Q which may be defined as 2π times the ratio of the mean energy stored in the resonator to the energy lost per cycle. Thus

$$Q = \omega U/W \tag{71}$$

where U is the stored energy and W the power dissipated. To calculate the Q of the resonator we must therefore first deter-

mine the fields inside it and then find by integration the energy of the field and the flow into the walls. Taking the scalar product of the first of Maxwell's equations with $\mathbf{H}^*$ and the conjugate of the second with $\mathbf{E}$ and subtracting we have

$$\mathbf{H}^* \cdot \operatorname{curl} \mathbf{E} - \mathbf{E} \cdot \operatorname{curl} \mathbf{H}^* = -g\mathbf{E}\cdot\mathbf{E}^* + j\omega\epsilon\mathbf{E}\cdot\mathbf{E}^* - j\omega\mu\mathbf{H}\cdot\mathbf{H}^*$$

$$= \operatorname{div} (\mathbf{E} \times \mathbf{H}^*). \tag{72}$$

Integrating over a volume bounded by a surface S,

$$\tfrac{1}{2}\iint (\mathbf{E}\times\mathbf{H}^*)\cdot\mathbf{dS} + \tfrac{1}{2}j\omega\iiint \mu\mathbf{H}\cdot\mathbf{H}^* dv - \tfrac{1}{2}j\omega\iiint \epsilon\mathbf{E}\cdot\mathbf{E}^* dv$$

$$+ \tfrac{1}{2}\iiint g\mathbf{E}\cdot\mathbf{E}^* dv = 0. \tag{73}$$

The real part of the complex Poynting vector $\frac{1}{2}\mathbf{E}\times\mathbf{H}^*$ gives the mean power crossing unit area. In a uniform static field of strength E_0, the electric energy density is $\frac{1}{2}\epsilon E_0^2$; in an harmonic field $\mathbf{E}$, $\mathbf{H}$ (x, y, z) exp j $(\omega t + \alpha)$, the mean electric and magnetic energy densities are $\frac{1}{4}\epsilon\mathbf{E}\cdot\mathbf{E}^*$ and $\frac{1}{4}\mu\mathbf{H}\cdot\mathbf{H}^*$. If the conductivity of the dielectric is zero, the last term in equation (73) representing the energy lost in the dielectric, is zero. The first term is equal to the complex flow of power across the surface S. If the surface is perfectly conducting, the vector product of $\mathbf{E}$ and $\mathbf{H}$ has no component normal to the surface.

Consequently

$$\tfrac{1}{2}j\omega\iiint \mu\mathbf{H}\cdot\mathbf{H}^* dv - \tfrac{1}{2}j\omega\iiint \epsilon\mathbf{E}\cdot\mathbf{E}^* dv = 0 \tag{74}$$

and it follows that the mean values of the oscillating electric and magnetic energies are equal. The total energy stored in a

resonator is therefore twice the mean magnetic energy and is given by

$$U = \tfrac{1}{2} \iiint \mu \mathbf{H}\cdot\mathbf{H}^* \, dv. \tag{75}$$

The mean flow of energy W into the walls is the real part of $\frac{1}{2}\iint E_t H_t^* dS$ where E_t is the small tangential component of electric field and dS is an element of area of the wall. Now

$$E_t = \mathsf{R}(1 + j)\, H_t, \tag{76}$$

therefore

$$W = \frac{\mathsf{R}}{2} \iint H_t H_t^* dS \tag{77}$$

$$= \frac{\mathsf{R}}{2} \iint \mathbf{H}\cdot\mathbf{H}^* dS.$$

Hence we have for the Q, assuming that the loss of energy is too small to appreciably affect the field distribution,

$$Q = \frac{\omega U}{W} = \frac{\mu\omega \iiint \mathbf{H}\cdot\mathbf{H}^* dv}{\mathsf{R} \iint \mathbf{H}\cdot\mathbf{H}^* dS} \tag{78}$$

where μ is the permeability of the dielectric and R the intrinsic resistance of the conductor. If we assume that dielectric and conductor have equal permeabilities, the formula becomes

$$Q = \frac{2 \iiint \mathbf{H}\cdot\mathbf{H}^* dv}{\delta \iiint \mathbf{H}\cdot\mathbf{H}^* dS} \tag{79}$$

where δ is the skin depth.

An idea can be obtained of the order of magnitude of Q's by assuming a uniform distribution of magnetic energy. A better approximation should result from halving the value obtained as magnetic fields tend to be strongest near the conducting surface.

Then

$$Q \doteqdot \frac{V}{\delta S} \tag{80}$$

where V is the volume and S the surface area. If the highest Q is to be obtained, the largest ratio of volume to surface area should be aimed at for the losses occur at the surface whereas the energy is stored throughout the volume. As the simple formula is approximate one cannot conclude that the resonator of highest Q for a given wavelength is spherical in shape. However the sphere is much better than either the cylinder or cube.

The Q of cylindrical resonators may be calculated from the following dimensionless formulae derived from equations (67), (69) and (79). The quantity $Q\delta/\lambda$ is a function of mode and shape only and is of the order of magnitude unity. In TM modes,

$$\frac{Q\delta}{\lambda} = \frac{\left\{\left(\frac{x_{lm}}{2\pi}\right)^2 + \frac{n^2}{4}\left(\frac{a}{h}\right)^2\right\}^{\frac{1}{2}}}{1 + \frac{2a}{h}} \qquad n \neq 0$$

$$= \frac{\left\{\left(\frac{x_{lm}}{2\pi}\right)^2 + \frac{n^2}{4}\left(\frac{a}{h}\right)^2\right\}^{\frac{1}{2}}}{1 + \frac{a}{h}} \qquad n = 0 \tag{81}$$

and for TE modes,

$$\frac{Q\delta}{\lambda} = \frac{4\pi^2\left\{\left(\frac{x_{lm}}{2\pi}\right)^2 + \frac{n^2}{4}\left(\frac{a}{h}\right)^2\right\}^{3/2}\left\{1 - \left(\frac{l}{x_{lm}}\right)^2\right\}}{x^2{}_{lm} + 2n^2\pi^2\left(\frac{a}{h}\right)^3 + \frac{l^2n^2\pi^2}{x^2{}_{lm}}\left(\frac{a}{h}\right)^3\left(\frac{h}{a} - 2\right)}. \tag{82}$$

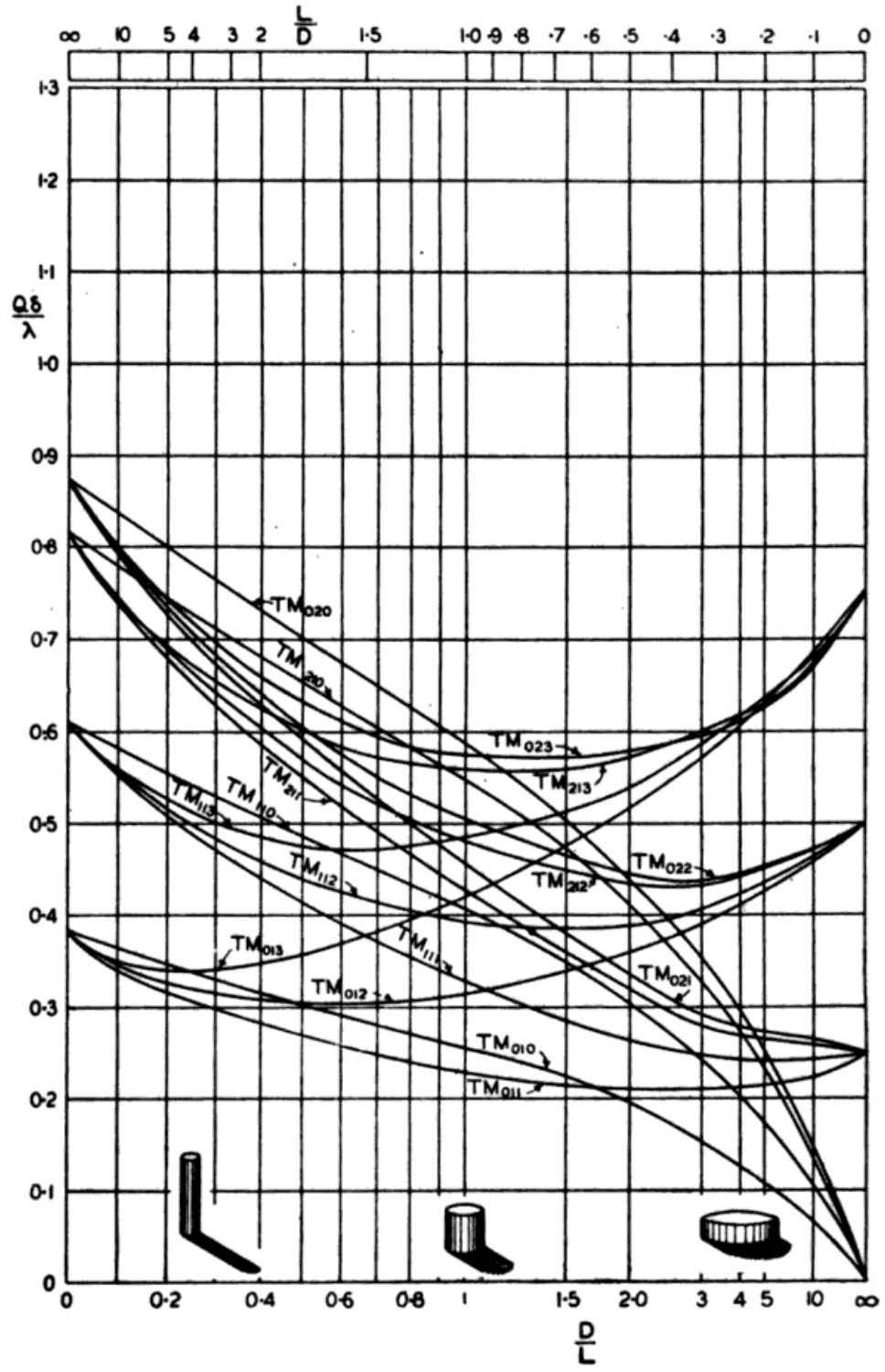


Figure 90.—$Q\delta/\lambda$ for TM modes in circular cylinders.

Equivalent Circuit

The behaviour of a resonant cavity in the neighbourhood of a resonance is reminiscent of the parallel resonant circuit with small losses and suggests the existence of a parallel combination of L, C and R which would accurately represent the resonator. To specify such a parallel circuit three quantities are necessary. These might be chosen as $(LC)^{-\frac{1}{2}}$, $R(C/L)^{\frac{1}{2}}$ and $(L/C)^{\frac{1}{2}}$ which are respectively the resonant frequency ω_0, Q and characteristic impedance. Now only two quantities are intrinsic in a resonator—resonant frequency and Q. Since there is nothing corresponding to characteristic impedance, none of the quantities L, C and R is defined. However by making one arbitrary assumption one may obtain an equivalent circuit which is then not unique. For example one may assume a value of L, C, R or $(L/C)^{\frac{1}{2}}$ or arbitrarily define the voltage or current associated with the circuit.

Cavity resonators, which are used for accelerating electron streams Fig 91, have a gap across which a voltage $V = V_0 \exp j(\omega t + \alpha)$ appears when electromagnetic power W

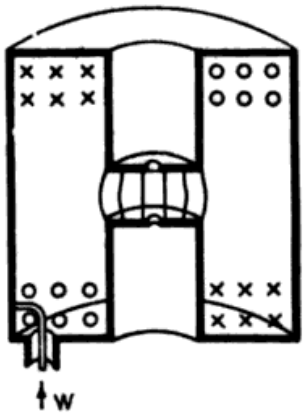


Figure 91.—Resonant cavity with gap for accelerating electrons.

is fed into the cavity. If the beam loads the cavity lightly, this power is absorbed in the walls. The suitability of the cavity for producing a high gap voltage from a given input power is measured by the shunt resistance R_{sh} which is defined so that

$$\frac{VV^*}{2R_{sh}} = W, \tag{83}$$

allowed for and this is usually satisfactory as the behaviour of a resonator is dominated by the mode whose resonant frequency is close to the exciting frequency.

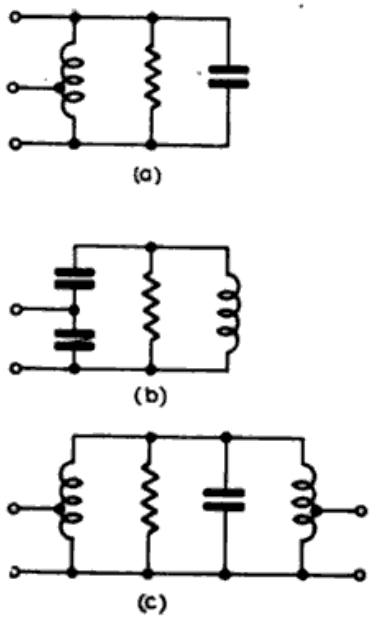


Figure 93.—Equivalent circuits for cavity resonators. (*a*) Cavity with magnetic coupling. (*b*) Cavity with electric coupling. (*c*) Cavity with two coupling loops.

BIBLIOGRAPHY

Proceedings at the Radiolocation Convention March-May 1946, *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946.

A. B. Bronwell and R. E. Beam, *Theory and Application of Microwaves*, McGraw-Hill, New York, 1947.

L. G. H. Huxley, *Survey of the Principles and Practice of Wave Guides*, Cambridge University Press, 1947.

W. Jackson, *High Frequency Transmission Lines*. Methuen, London, 1945.

R. L. Lamont, *Waveguides*. Methuen, London, 1942.

S. Ramo and J. R. Whinnery, *Fields and Waves in Modern Radio*. Wiley, New York, 1944.

R. I. Sarbacher and W. A. Edson, *Hyper and Ultrahigh Frequency Engineering*. Wiley, New York, 1943.

S. A. Schelkunoff, *Electromagnetic Waves*. Van Nostrand, New York, 1943.

J. C. Slater, *Microwave Transmission*. McGraw-Hill, New York, 1942.

CHAPTER VII

TRANSMISSION LINE AND RESONATOR TECHNIQUES

In the previous chapter the theory of waveguides and cavity resonators has been discussed. It is now necessary to relate that theory to practical applications, and to discuss in some detail the more significant of these applications. In actual fact, more than the direct transfer of theory to practical equipment (i.e. working models) is involved; such aspects as economic considerations, ease of operation, wide utility etc., must also be considered. It is with these aspects as well as the application of theory that the following discussion will be concerned.

When a particular system is to be described, it will be designated as either coaxial line or waveguide, these being the only two which are of concern in this chapter.

1. Feeder Systems in General

In a radar set, the feeder system connects the transmitter to the aerial and the aerial to the receiver and includes the auxiliary parts such as TR and RT switches.

The main requirements of such a system are that the RF power be transmitted with negligible or small losses, and with freedom from electrical breakdown. In addition, as the radar set may be required to work on any one of a number of frequencies in a given frequency band, the feeder system and its associated parts must function equally well over the specified band. This eliminates the necessity for readjustment and tuning of the apparatus every time a change of frequency is made. In practice however, it may not be possible to obtain uniformly good performance over the entire band, due to the frequency-selective elements employed, and so a limit is set for the worst performance that can be tolerated.

In case (*b*) where $Z_o \neq Z$, a matching device whose nature is of no concern here must be inserted as shown. At *A* the impedance looking towards the load, assuming of course that the matching device has been correctly adjusted, is equal to Z_o. Then, if the matching device is dissipationless, there is a maximum transfer of energy to the load. The matching

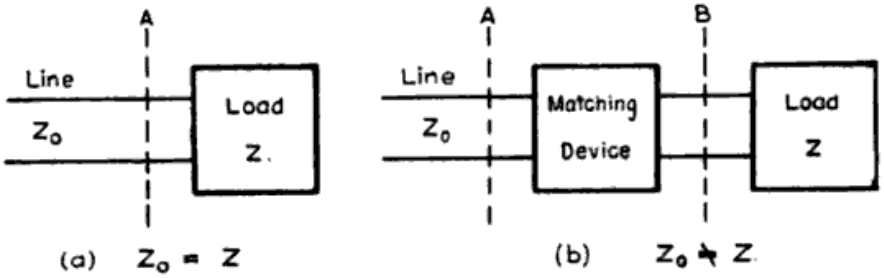


Figure 94.—Various line terminations showing impedance and matching relationships.

device thus transforms the load impedance Z to the line impedance Z_o, and hence is often termed a transformer.

In matching, it is first necessary to know with some precision the impedance of the load. The most convenient method at very high frequencies is to make this measurement by determining the standing waves produced when this load is used to terminate a line of known impedance. The relations between impedance, voltage standing wave ratio S, and the physical displacement of the voltage node from the load have been discussed in the previous chapter. Although the equipment used will be discussed fairly fully in the section on measurements, it is advisable to give at this stage at least an idea of its nature. In the case of coaxial transmission lines, the necessary information may be obtained by a pick-up probe which projects a small distance into a narrow longitudinal slot in the line, and which may be moved along it. Fig. 95 shows the complete "set-up" with detecting crystal and meter.

The arrangement of the slot does not affect the field distribution inside the line, because for the *TEM* mode, which has only axial current flow, it intercepts no current, assuming

that the slot width is small in comparison with the main dimensions. Plate X (*a*) shows a photograph of this type of equipment.

Once the load impedance has been determined it is convenient to work in terms of normalised impedance. This is

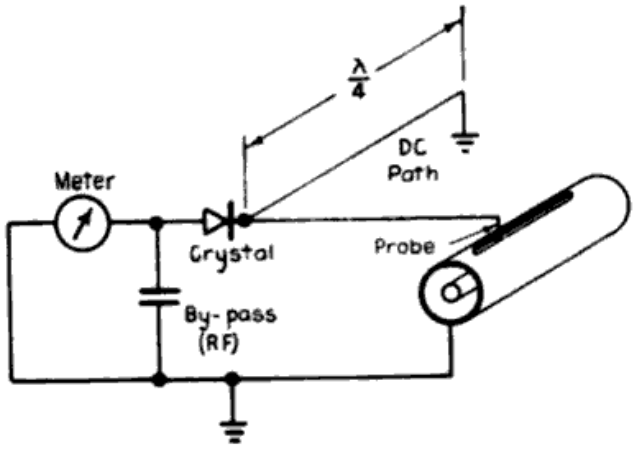


Figure 95.—Impedance measuring equipment. Electrical circuit details.

obtained by dividing the impedance by the characteristic impedance. It is equally suitable to work in terms of admittances rather than impedances, the actual case usually dictating the choice. When the normalised impedance or admittance is known it is necessary to decide how the transformation to the required value is to be made. This is easily done by first plotting the load impedance on either the Smith or Kennelly charts.

Consider for example the point P in Fig. 96 representing an impedance $0{\cdot}6 - j0{\cdot}7$ (normalised values) plotted on the Smith chart. To transform this value to that of $1 + j0$, point O, there are several possible devices that may be used :

(*a*) A reactance of magnitude $+ j0{\cdot}7$ may be added at the load. This transforms point P to S along the path shown, giving a resistive value 0·6. Then, by means of a quarter-wave section, this value is transformed to $1{\cdot}0 + j0$. [If Z is the normalised characteristic impedance of the $\lambda/4$ section, then $Z^2 = (0{\cdot}6 \times 1{\cdot}0)$. Generally $Z^2 = (Z_1 Z_2)$. For further explanation see standard reference books.]

The determination of the actual dimensions is a little more difficult in this case however, as the transformation follows the curve $P'V$ in Fig. 96. [See Fig. 97 (*b*).]

So far, the discussion has been limited to the matching of

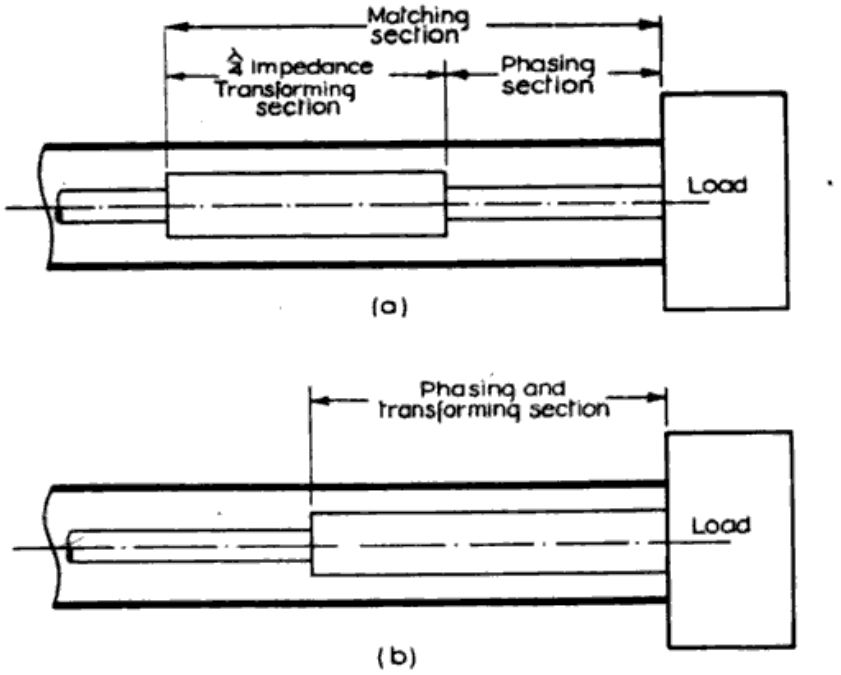


Figure 97.—Matching methods. Physical arrangement of matching devices.

particular elements and the methods that may be employed. It is also desirable to have a unit which may be used to match any arbitrary element that may be chosen. This immediately presupposes variable elements, thus eliminating the possibility of the normal $\lambda/4$ sections. Such units have provision for adding shunt reactances at various points of the line.

A convenient unit has two reactance elements spaced $\lambda/8$ apart and placed so that they are effectively in parallel with the transmission line. Fig. 98 shows suitable elements for coaxial lines. The actual function of this transformer is best illustrated by reference to the Smith chart (see Fig. 99), which may be used for admittances by reading off values as G and B, instead of R and X, as in the case of impedances. The admittance is the reciprocal of the complex

quantity $R + jX$, and this transformation may be carried out on the Smith chart by rotating the impedance point 180 geometrical degrees about the centre and reading the real and imaginary components obtained as conductance and susceptance respectively.

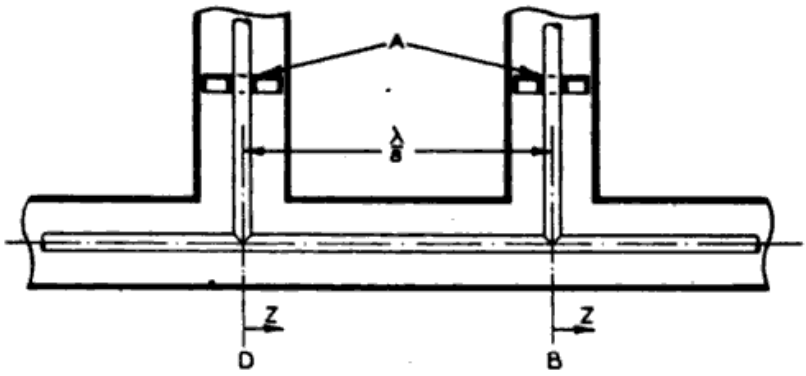


Figure 98.—Two stub matching transformer. Physical arrangement showing stubs and their movable shorting plungers *A*.

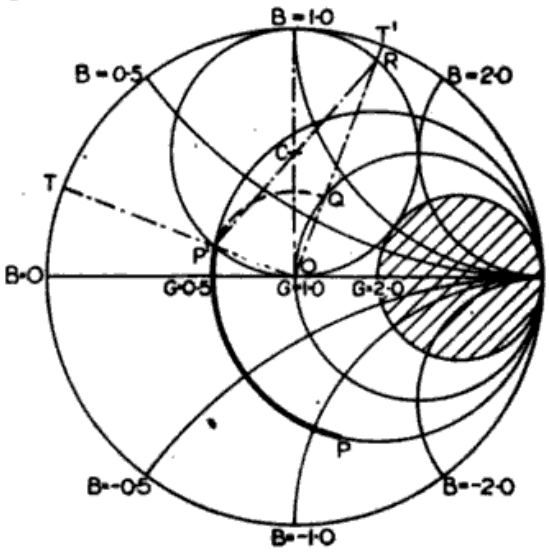


Figure 99.—Tuning range of two stub matching transformer. The transformer may be used to match any impedances except those which lie within the shaded circle.

Assume that point P corresponds to the load admittance which is to be matched to the line admittance (point O). The further assumption is made that the reactance elements may be adjusted to give any reactance between $+\infty$ and $-\infty$.

quencies above and below the design frequency the transformer is not $\lambda/4$, and so introduces a reactance which in turn means a reflection of power. This is shown in Fig. 101 by points Z_H and Z_S.

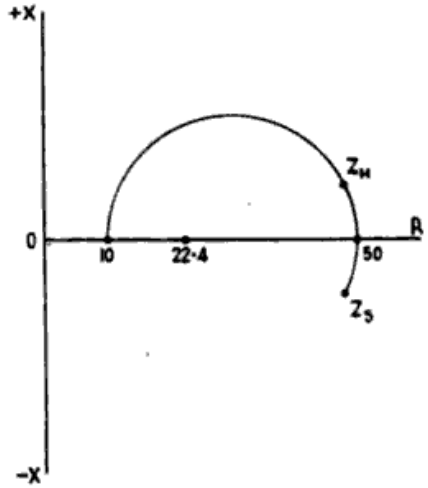


Figure 101.—Impedance variation with frequency for a high ratio quarter-wave transformer.

Consider now the transformation from say $40 + j0$ to $50 + j0$

$$Z = (40 \times 50)^{\frac{1}{2}}$$
$$= 44{\cdot}6 \text{ ohms.}$$

This transformation is plotted to the same scale in Fig. 102.

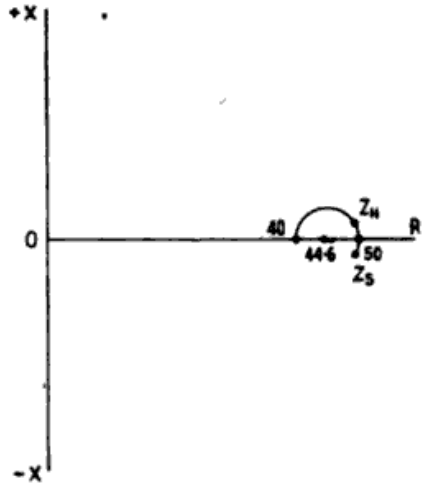


Figure 102.—Impedance variation with frequency for a low ratio quarter-wave transformer.

For the same percentage change of frequency as in Fig. 101, the corresponding points Z_H and Z_S are closer to the required

In practice, most equipment is designed assuming perfect conductors, and then with the magnitudes of the current flow thus obtained and a knowledge of the resistive properties of the materials involved, the actual losses are determined.

At microwave frequencies, current flow in a conductor is determined principally by " skin effect," a discussion of which has already been given in the previous chapter. If δ, the depth of penetration, is defined as the depth below the surface of a conductor at which the current density drops to $1/e$ of its surface value,

$$\delta = 0{\cdot}029\left(\frac{\lambda}{\mu_r g}\right)^{\frac{1}{2}}$$

where g is the conductivity of the conductor in mhos per centimetre, μ_r is the relative permeability (μ/μ_o), and λ is the wavelength in free space in centimetres.

Fig. 106 gives values of δ for different materials for wavelengths up to 30 centimetres. As δ becomes smaller, the current flow becomes greater at the surface of the conductor,

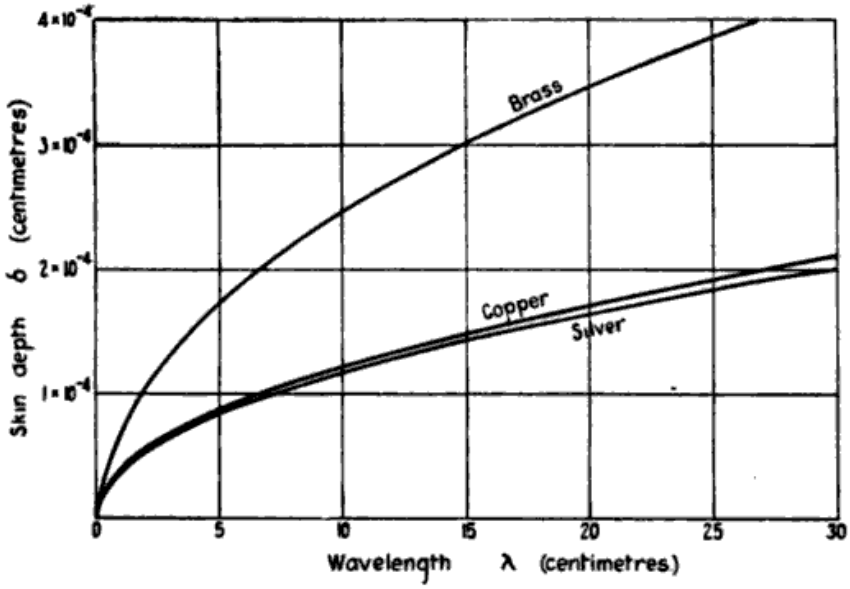


Figure 106.—The depth of penetration (δ) of currents in various conductors at very short wavelengths.

and it is because of the excessively high surface current densities that there are considerable losses.

The actual surface of the material is of some importance, depending particularly on the application. Very little quan-

decreased by a factor of safety to give a suitable working potential difference. This will make allowance for resonant voltage step-ups which occur in matching devices, chokes, etc., as well as preclude the possibility of breakdown due to varying atmospheric conditions. The breakdown stress is proportional to the square root of the pressure, a fact of considerable importance in equipment designed for use at high altitude.

The power that can be transmitted in a feeder is given by

$$P = 6{\cdot}63\ E^2.\ a.b.\ \frac{\lambda}{\lambda_g}\ 10^{-4}\ \text{watts}$$

for a waveguide[6] (TE_{10} mode) and

$$P = \frac{E^2 r_1^2 Z_0}{7200}\ \text{watts}$$

for a coaxial line where E is the peak electric field strength.

At $\lambda = 10$ centimetres and taking $1{\cdot}5 \times 10^4$ volts per centimetre as the maximum-permissible electric stress, a waveguide 3 inches $\times$ $1\frac{1}{2}$ inches will carry 2·4 megawatts of power, whereas a 50 ohm coaxial line, $1\frac{5}{8}$ inches outer conductor, will only carry 1·0 megawatts.

In practice, it is usual to determine the factor of safety on a power basis and a value of 0·2 to 0·25 of the maximum value is commonly used.

5. Design Aspects

The first point to be considered in designing feeders and associated equipment is the type of feeder to be used. Twin wire lines are automatically excluded because of their high radiation at microwave frequencies. Coaxial lines may be used at frequencies as high as 3,000 megacycles per second but attenuation is fairly severe at the higher limit. In addition, the physical dimensions must be kept sufficiently small to prevent a higher transmission mode existing in the line. It should be emphasised that in all transmission systems there should be only one mode of energy propagation. If two or more exist, the field patterns become complicated and considerable difficulty would exist in designing equipment to

[6] See Chapter VI, equation (29).

has been fixed, it is necessary to limit the actual dimensions, so that the mean circumference $\pi(r_1 + r_2)$ is less than λ. If this is not done the TE_{11} mode is capable of being propagated.

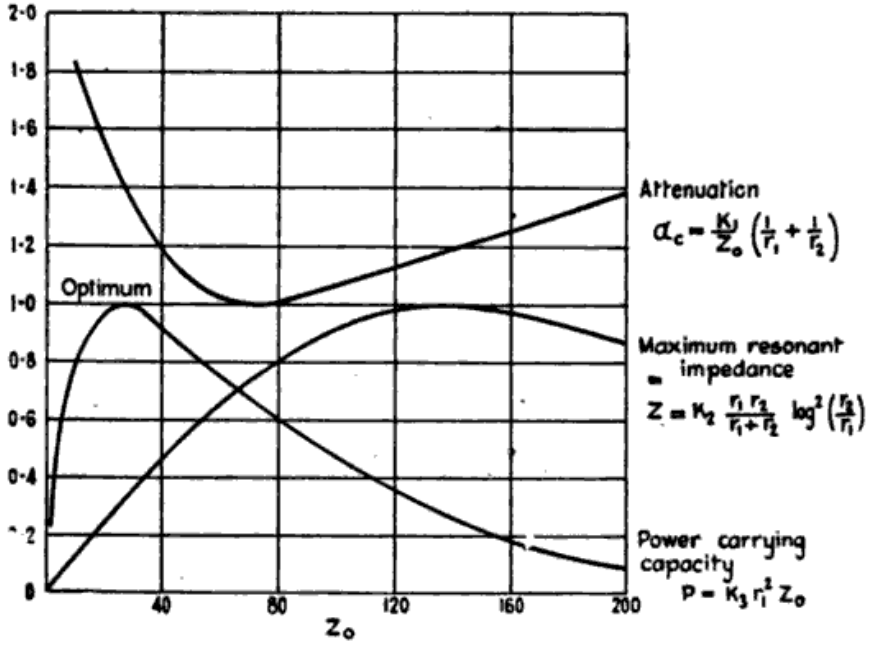


Figure 111.—Characteristics of coaxial lines as a function of Z_0 for a fixed value of r_2. Curves are referred to an optimum value of unity.

Leaving aside connection by cables it is apparent that for high power work, rigid elements must be considered, which for convenience do not normally exceed about three feet in length.

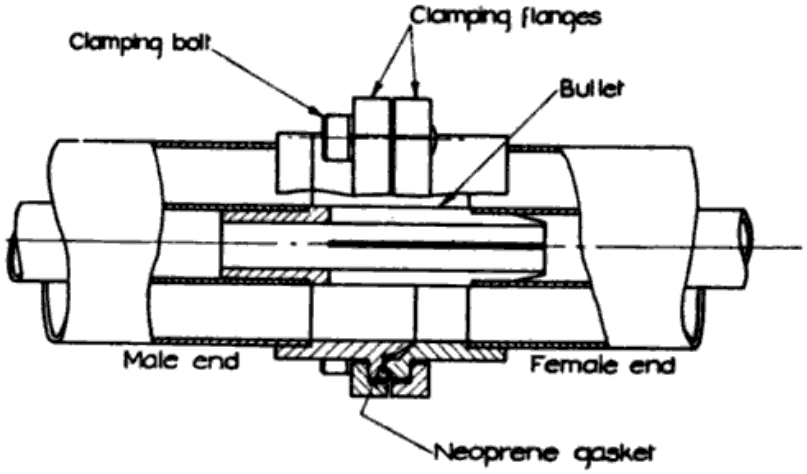


Figure 112.—Coaxial line joint for a 1⅝ in. O.D. line.

The problem is then how to make joints, bends, allow rotating motion, etc.

The determination of the fields in a waveguide is treated in Chapter VI.

The selection of the dimensions of a waveguide for a particular wavelength is governed mainly by the cut-off or critical frequency, the power capacity, and the attenuation desired. With rectangular guides, the most commonly used mode, which gives the smallest physical dimension is the TE_{10}. This has a field distribution as shown in Fig. 116. The current

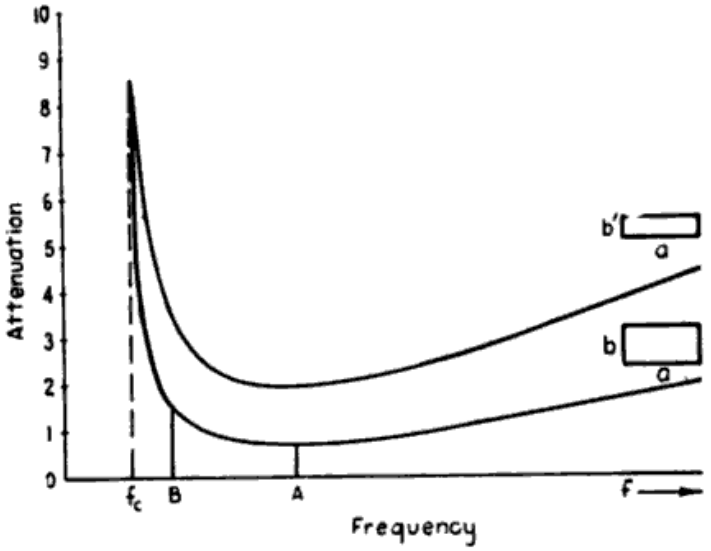


Figure 117.—Variation of attenuation with frequency for waveguides having different dimensions b. (TE_{10} mode.)

flow is shown on the surface of the guide and the field distribution across the cross section.

Fig. 117 shows the variation of attenuation in a waveguide with frequency and the ratio a/b for a TE_{10} mode. Although point A corresponds to minimum attenuation the dimension a is usually chosen to give an operating frequency B approximately 1·5 times f_c. Greater values of B allow the possibility of higher order modes being propagated and must be avoided even though the attenuation is slightly increased.

In the case of the TE_{10} mode where the critical frequency is independent of the dimension b, this dimension must be chosen from a consideration of the permissible attenuation and the power to be handled by the waveguide. The question

Twists may be made in waveguides provided that the guide is not deformed unduly in cross section. For small reflection, the length of the twist should be of the order of one wavelength, or more. Flexible guide may also be used in certain applications; however the attenuation is somewhat greater than for solid guide, and some reflection of power is present. Flexible guide may be made in a number of ways—in bellows form with each fold a half wavelength long or wound from strip as in the construction for flexible metallic piping.

A taper may be used in a waveguide to transform from one mode to another; for example, from a TE_{11} mode in a round guide to a TE_{10} in a rectangular guide. This will be seen to be a transformation from the lowest mode in one guide to the lowest in another. Transformation between higher modes usually introduces unwanted modes which then necessitate filter devices. The reflections set up at such a taper are a function of the length of the taper.

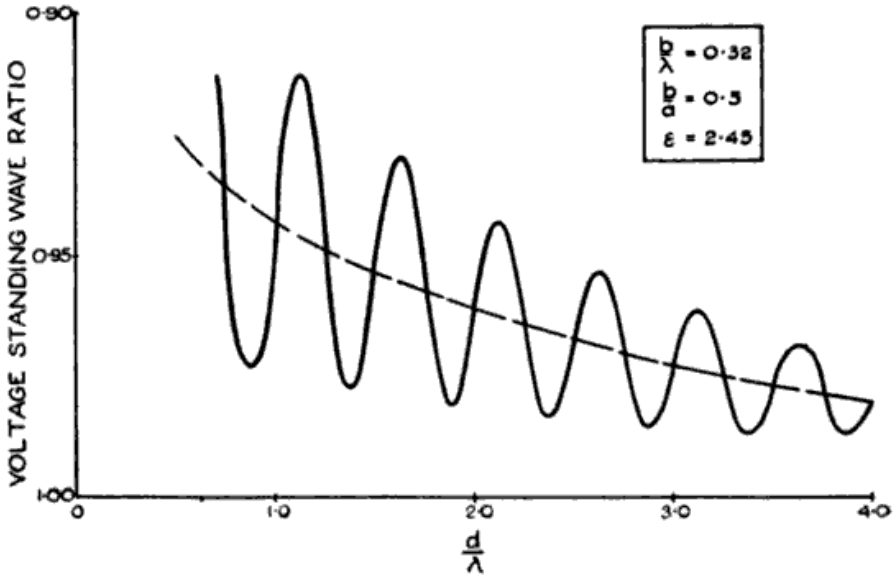
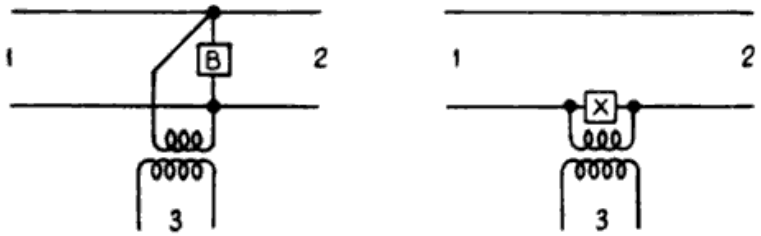


Figure 122.—Reflections from a tapered dielectric section of length d inserted in a waveguide; taper in E-plane.

A change in dielectric in a guide may be conveniently made by taper sections, starting from zero thickness and increasing to full guide section in a distance d. Fig. 122 shows the theoretical reflection set up from such a taper. The transition could of course have been made by means of quarter-wave

devices to filters of all types will be immediately apparent.

T-junctions in waveguides form an important class of elements, both as dividing networks and as matching elements, and they may be treated in much the same way as conventional six terminal networks as far as impedance properties are con-



(a) H Junction (b) E Junction

Figure 125.—Equivalent circuits of T-junctions in waveguides.

cerned. With rectangular guides, junctions may be made in two ways, either in the broad or narrow sides. The impedance relations in either case are obtained by setting up the wave equations at the junction and solving for the correct boundary conditions. This involves higher order modes which, as in the case of irises are rapidly attenuated at a distance from the junction. With a TE_{10} wave, a junction in the broad side of the guide produces a predominantly series loading to the main guide and is called an E-junction. A junction in the broad side produces predominantly shunt loading and is termed an H-junction. This is not necessarily true for other main modes of propagation.

The T-junction can best be represented as shown in Fig. 125, where 1 and 2 are the main arms, and 3 is the side arm. In either the H or E types, there is a common susceptance or reactance at the junction and the side arm is connected through a step down transformer. The effect, besides giving a step down of impedances from the side arm to the main arm, produces a phase shift at the junction which must be considered in the design. To overcome the step down transformer action, an iris of suitable dimensions or some matching section must be inserted in the side arm guide. If this is not done, a correct termination in the side arm will not appear as a

correct termination in the main guide when a shorting plunger is placed in the main guide on the opposite side of the junction, and at a correct distance from it.

The impedance or admittance relationships may easily be determined experimentally by placing known impedances at two points, and measuring the third impedance. This simple method will give quite reasonable results.

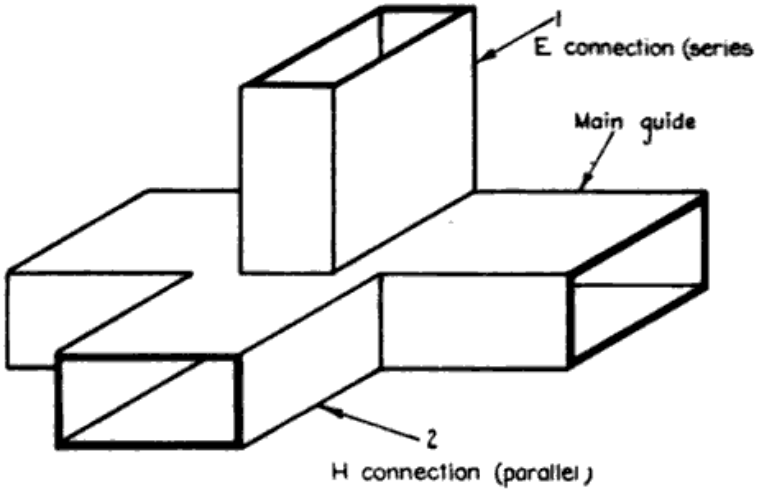


Figure 126.—" Magic tee " junction in a rectangular waveguide, showing the side *E* and *H* connections.

By placing a shorting plunger in the side waveguide of a junction, varying values of reactance or susceptance may be introduced into the main line. Thus, very convenient matching elements may be made.

By adding a second side waveguide at the junction an eight terminal network is produced which has also certain impedance symmetries. One special case of interest is that of the " magic tee," in which there is one side connection to the broad side and one to the narrow side of the main guide. Fig. 126 shows the general arrangement.

Consider now a TE_{10} wave incident at the junction in guide 2. The magnetic field will couple into both arms of the main guide and excite waves in them which are in phase. No wave will be excited in guide 1 because neither the magnetic nor the electric fields present will couple to it to produce a TE_{10} wave. If now a TE_{10} wave in guide 1 is incident at the

guide to a TM_{01} mode in a circular guide. This may be done in several ways, the criterion of performance being that reflections should be negligible and that only the TM_{01} mode exists in the circular guide. Fig. 129 indicates one way in which it may be done.

From previous remarks the stub will be seen to be effectively in series with both guides.

The resonant ring is $\lambda_{TE_{11}}/4$ from the centre of the junction and presents a high series impedance to the TE_{11} mode. The main stub being $\lambda_{TM_{01}}/2$, (d_2), presents a low series impedance to the TM_{01} mode. Thus the TE_{11} mode in the circular guide

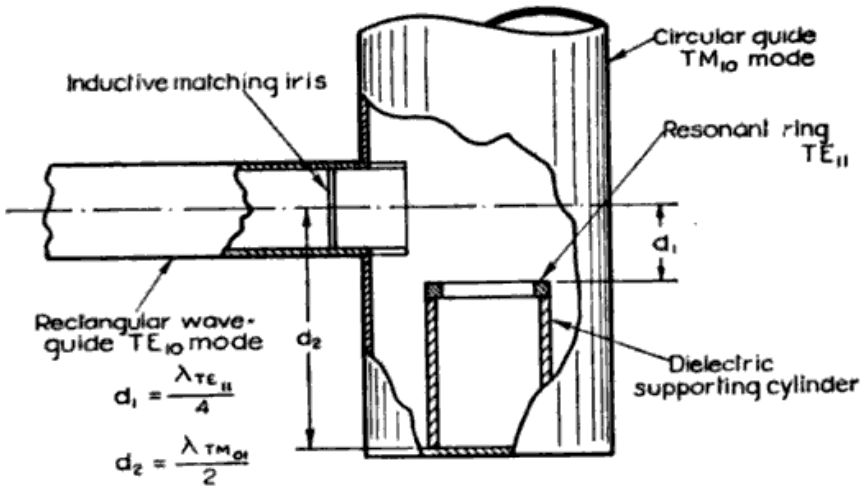


Figure 129.—TE_{10}—TM_{01} transformer for transforming a TE_{10} wave in a rectangular guide to a TM_{01} wave in a circular guide.

is inhibited and the TM_{01} encouraged. In this particular transformation an inductive matching iris is necessary in the rectangular guide to provide correct matching into the transformer. If the device is dissipationless, then by the reciprocity theorem, the transformation from TM_{01} to TE_{10} may be made with this same device by reversing the power flow.

For transforming from coaxial line to waveguide, a probe or loop from the coaxial line must be inserted in the waveguide. For a matched condition to be obtained, the probe or loop

must be resonant; this means the probe must be approximately a quarter of a wavelength, and the loop approximately a half wavelength long. Two typical couplings are illustrated in Fig. 130. It will be noticed that in the high power feed (*b*),

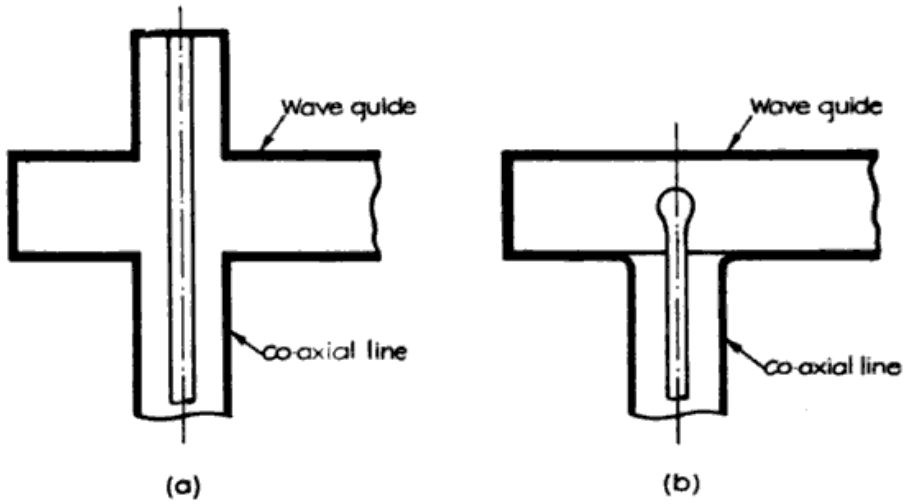


Figure 130.—Coaxial to waveguide couplings.

edges are carefully rounded to prevent high localised electric stresses.

6. Cavity Resonators

The uses of cavity resonators are numerous, klystrons, magnetrons, wavemeters and filters being but a few of the many applications. Although in theory resonators may have any shape, in practice they are usually restricted to about three main forms—circular cylinders, coaxial cylinders and rectangular prisms. Because of ease of fabrication, the circular cylindrical types are the most common. The resonant frequencies of these types of resonator have been given in the previous chapter.

The coupling to a cavity can be considered as a means of forming a flux distribution inside the cavity. As described in the previous chapter, this may take the form of either loops, probes or windows, depending on the feed system employed. Now, the nature and position of the feed influence the type of mode excited, and by careful design it is

to represent a resonator in terms of a series loss resistance and a series inductance and capacity.

Coupling to a cavity is often represented by an ideal transformer with an impedance transformation ratio of K^2. Fig. 133

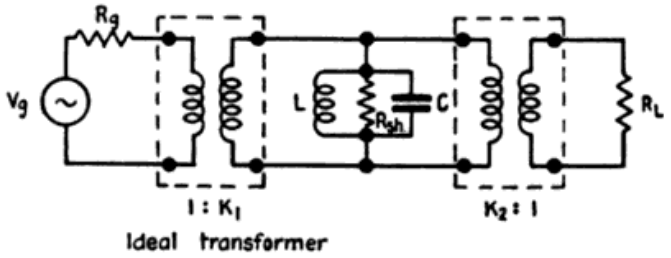


Figure 133.—Equivalent circuit of a resonator coupled to a generator V_g and to a load R_L.

illustrates a case where a cavity is employed as a band pass filter coupling a generator to a load as shown. The circuit may be simplified as shown in Fig. 134 which may then be

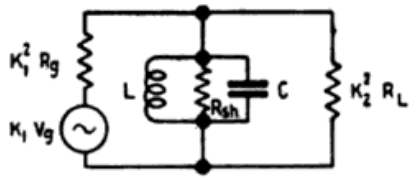


Figure 134.—Simplified form of Fig. 133.

analysed to determine its behaviour. As most cavities are usually employed in a resonant condition the circuit may be further simplified as shown in Fig. 135. In this circuit, the input and output couplings K_1 and K_2 are entirely arbitrary.

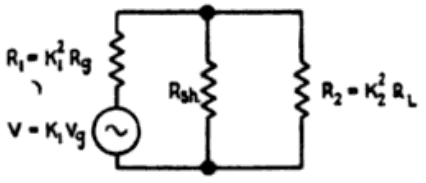


Figure 135.—Simplified circuit of Fig. 133 at the resonant frequency of the cavity.

They are often chosen to make the input impedance match that of the generator.

in common use in this laboratory will be described. Hence, the following remarks must be taken as indicative only of some of the methods which may be employed.

Frequency and Wavelength

The question at once arises, which of the two is to be employed. If we had an exact knowledge of the physical properties of various media it would be immaterial which were used. This is not so, however, and for exact work it is preferable to decide on one or the other. Frequency being the more fundamental, and being independent of the medium, is the desirable choice. However, in most laboratory measurements where accuracy is not paramount it is immaterial which unit is used, a value of 3×10^{10} centimetres per second for the velocity of wave propagation giving sufficient accuracy for the transformation.

Coaxial lines may be used for measuring wavelength by using a short length of line with a movable short circuiting plunger. By measuring the distance between two successive resonance points the wavelength may be determined. A typical example is shown in Fig. 137.

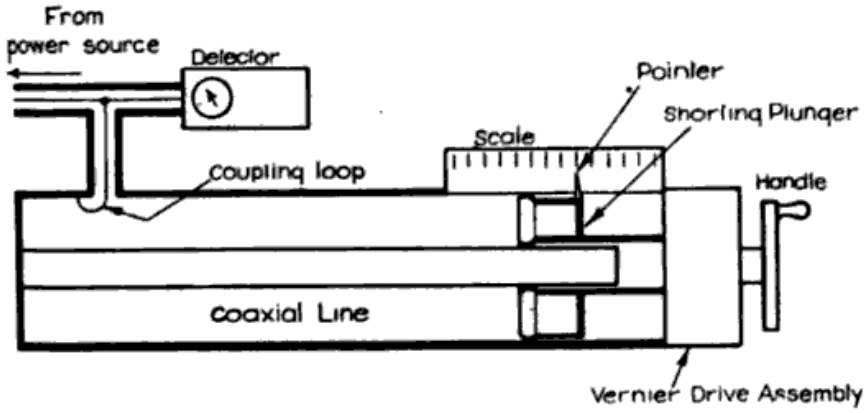


Figure 137.—Coaxial type wavemeter. Resonance is indicated by a dip on the detector meter.

In the particular device shown, resonance is indicated by a dip in the crystal current reading. The corrections for field distortion at the loop etc., are very small, and in many cases a direct reading of the distance will give sufficient accuracy.

Mention has already been made in section 6 of the method of tuning a cavity resonator. Fig. 131 shows a semi-coaxial type which may very conveniently be used as a wavemeter. Alternatively the end plate of a hollow cavity may be moved in or out to provide means of measuring wavelength. Such devices must be calibrated either by comparison with a coaxial type of instrument, or by comparison with the higher harmonics of a standard crystal oscillator. A crystal and meter are also coupled into the cavity in order to indicate the resonance point. Whilst the coaxial type gives a fundamental length measurement, it will be seen that the cavity type is only a comparison method and hence requires careful calibration if accuracy is to be achieved. Fixed wavelength resonators may be made and their resonant wavelengths determined by a length measurement. These are then useful as reference standards to provide check points in the various wavelength bands. Bracewell[7] describes a standard cavity of the type mentioned and also the construction and calibration of a cavity type wavemeter against a standard crystal.

In certain cases it is permissible to connect the wavemeter direct to the oscillator or signal source. For high accuracy however, this is not good practice, for the following reason. Most signal sources, particularly at microwave frequencies, contain only the oscillator with an output connection directly from it. As these oscillators do not supply much power, it is necessary to couple tightly to them which then means that a fairly large external impedance is reflected into the oscillatory circuit, thus in part determining the frequency of oscillation. If the load changes through connecting and disconnecting the wavemeter, then the oscillator frequency may change. Tuning through resonance on the wavemeter will cause large impedance changes and hence may exert a considerable effect on the frequency. In practice it is best to isolate the oscillator with at least 10 decibels attenuation. Indeed it is preferable to use a directional coupler with a ratio of 100 to 1 to enable

[7] R. N. Bracewell, "L-band standard cavity," C.S.I.R. Radiophysics Laboratory, Report TI 134/2, September, 1944; and "L-band cavity wavemeter," C.S.I.R. Radiophysics Laboratory, Report TI 134/1, September, 1944.

exercised in their use. Thermistors have a tiny bead made out of a mixture of metallic oxides (Mn_2O_3, SiO) and have a negative coefficient of resistance. Their sensitivity is equivalent to that of the bolometer, and at the same time they are much more robust.

The actual element to be used must be mounted in a suitable holder to enable it to absorb all the power and to enable DC connections to be made to it.

Such a device is shown in Fig. 139 with a matching screw to take account of variations in individual thermistors. It is fairly tolerant to frequency changes and requires few or no correcting factors.

A balanced bridge is used with this instrument to determine the power. As a rule, the instrument is calibrated on DC or 50 cycles AC and the effective range is from about 10 microwatts to 4 to 5 milliwatts.

A typical circuit is shown in Fig. 140.

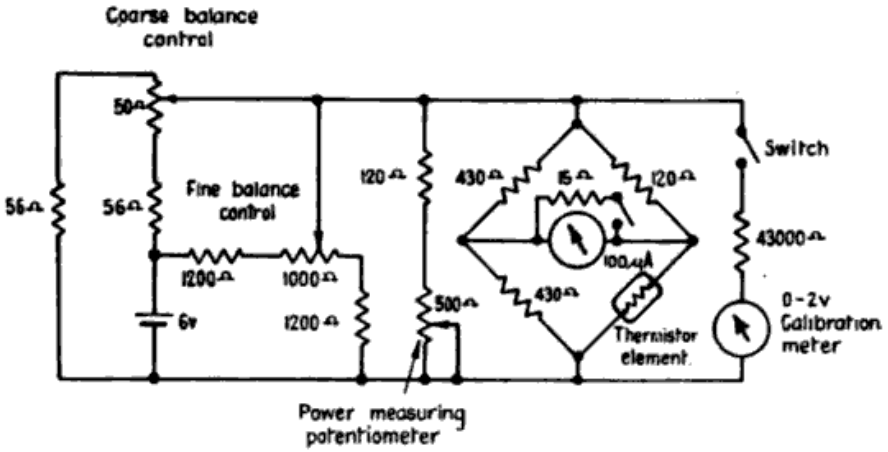


Figure 140.—Thermistor bridge circuit used to measure powers of 100 microwatts to 5 milliwatts.

For measuring high powers on a low power instrument a directional coupler (section 5*b*) or a known attenuator must be used. The attenuation may be either fixed (lossy cable or a lossy element in the feeder) or may be of a variable type. The variable types make use of the characteristics of a waveguide operated at wavelengths longer than the critical value.

The propagation constant in a waveguide is given by equation 17 of Chapter VI.

$$\Gamma^2 = \left(\frac{2\pi}{\lambda_c}\right)^2 - \omega^2\epsilon\mu .$$

If $(2\pi/\lambda_c)^2 < \omega^2\epsilon\mu$, Γ is imaginary and so gives propagation with a real velocity and no attenuation. If, however, $(2\pi/\lambda_c)^2 > \omega^2\epsilon\mu$, Γ is real and represents attenuation.

Taking the latter case, it can be shown that

$$\alpha = \frac{2\pi}{\lambda_c}\sqrt{1 - \left(\frac{\lambda_c}{\lambda}\right)^2} \text{ nepers per metre.}$$

This gives an attenuation device which may easily and correctly be calibrated by a length measurement and which moreover has the advantage of providing a linear variation when cali-

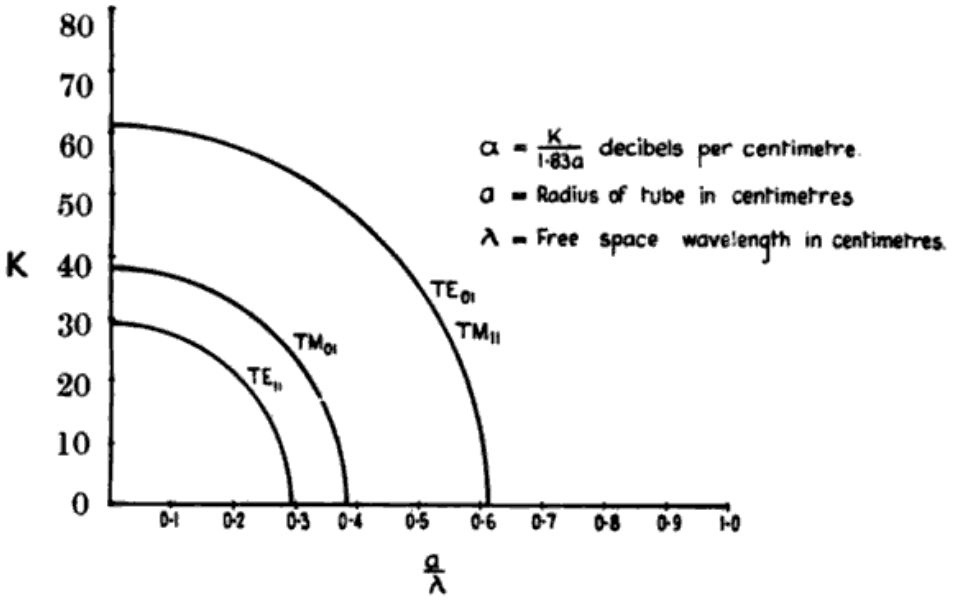


Figure 141.—Attenuation in a circular waveguide which is operated below its critical frequency.

brated in decibels. The most commonly used excitation is that in circular guides of TE_{11} or TM_{01} modes. In Fig. 141 a graphical means is given of determining the attenuation for any of several modes.

CHAPTER VIII

AERIALS

THE development of ultra high frequency and microwave techniques during the last few years has given rise to a corresponding development in aerial design. The tendency has been to produce either very narrow beams or beams with a predetermined directional characteristic. To some extent this has occurred through modification of long wave techniques, e.g. the use of arrays of dipoles etc., but, because of the fact that it is possible at these frequencies to make aerials with linear dimensions large with respect to the wavelength, the problem has also been approached from the point of view of geometrical optics, and this has resulted in aerials using paraboloids, lenses, etc. A further type of aerial has developed out of the obvious use of an open ended waveguide as a radiator, namely the flared waveguide or horn aerial.

The theory of directional aerials at these frequencies is very similar to the theory of diffraction in optics. For instance, an array of point sources is analogous to a diffraction grating, the radiated field from a paraboloid fed by a point source corresponds to the optical case of Fraunhofer diffraction at an aperture, and that of a sectoral horn arises from Fresnel diffraction at its aperture.

The various types of aerials have been broadly grouped into the following three categories for the purposes of discussion.

(*a*) Elementary radiators, of which the half wave dipole is one of the most important types.

(*b*) Arrays of elementary radiators.

(*c*) Continuous surface radiators. These are aerials in which the radiation is produced by a continuous current distribution in a conducting sheet or by a continuous field distribution across an aperture, where the dimensions of the

radiation to the average radiated power per unit solid angle

$$\text{i.e. Directivity} = \frac{4\pi\Phi_{max}}{\int\Phi d\Omega}$$

$$= \frac{4\pi}{\int\Phi^* d\Omega} \qquad (5)$$

where Ω is the solid angle.

The gain of an aerial is defined relative to a standard radiator, and is the ratio of the power which must be supplied to the standard to produce a given field at a given distance in the direction of maximum radiation, to that which must be supplied to the aerial to produce the same field at the same distance. The standard aerial may be a hypothetical spherical (isotropic) radiator, a half wave dipole, or a short current element (infinitesimal dipole). The factors relating these three definitions of gain may be easily found, since the gain, G_s, of an infinitesimal dipole relative to a spherical radiator is 1·5 and that of a half wave dipole is 1·64. For an aerial in which there are no ohmic losses, the gain relative to a spherical radiator is equal to the directivity. An example for which this is not so is that of a rhombic aerial in which some of the power supplied to the system is dissipated in a resistive load. Another example is the non-resonant array of section 4.

Except in a few cases, the term *gain* will be employed in this chapter even when strictly directivity is meant, since in most of the aerials described ohmic losses are negligible. The symbols G_s, G_c, G_d will be used for gain relative to a spherical, infinitesimal, or half wave radiator respectively.

Radiation Resistance

The radiation resistance of an aerial referred to a given point in the system is equal to the resistance which, if inserted at that point, would dissipate the same energy as is actually radiated by the system. If the current at the reference point is I_o, then

$$R_r = \frac{\text{Radiated power}}{I_o^2}. \qquad (6)$$

The concept of radiation resistance has its most important

to equal the energy actually removed and absorbed in the load. It may be shown that

$$A_o = \frac{G_s \lambda^2}{4\pi} \Phi^*(\theta,\phi). \tag{14}$$

$$\text{Hence } A_{o_{max}} = \frac{G_s \lambda^2}{4\pi} \tag{15}$$

when the aerial has optimum orientation. If the actual cross sectional area of the aerial equals the absorption cross section then

$$G_s = \frac{4\pi A}{\lambda^2}.$$

It will be seen in the next section that this is the gain of an aerial which, when transmitting, has a field distribution uniform in phase and amplitude across an area A.

Re-radiation Cross Section (A_r)

Since currents are induced in an aerial by the incident wave, re-radiation can take place, and the re-radiation cross section is defined as the area of the incident wave front from which energy would have to be absorbed to equal the energy re-radiated. It is often erroneously stated, on the basis of a simple circuit analogy, that for any matched aerial the received energy is equally distributed between load resistance and radiation resistance, and so the re-radiation cross section equals the absorption cross section. If this were so, the black body concept, so useful in thermo-dynamics would be invalid.

A more exact circuit analogy is given in Fig. 146. The infinite transmission line of characteristic impedance Z_o represents free space which guides the wave to the aerial represented by the impedance Z. The impedance looking into the line at AA' represents radiation resistance and equals $\frac{1}{2}Z_o$, and so for a matched aerial we must have $Z = \frac{1}{2}Z_o$. It can be shown that in this case, of the incident power P, $P/2$ is absorbed and $P/2$ "re-radiated," half of this being reflected and half transmitted. This case would correspond to an array of dipoles without any reflecting sheet, and so

Because of the high reactive component of the impedance of short current elements and because it is in general more convenient to use a resonant dipole, they are rarely used at ultra high frequency except where there is a restriction on size as in the case of detector probes for impedance measuring gear etc.

The Half Wave Dipole

A thin conductor one half wavelength long behaves as a resonant element, and if the current distribution along it is known the radiation field may be calculated by considering it as an aggregate of current elements. This treatment leads to the following results :

$$S(\theta,\phi) = \frac{\cos\left(\frac{\pi}{2}\sin\theta\right)}{\cos\theta} \tag{21}$$

where the dipole is centred at the origin of the coordinate system of Fig. 144 and lies along the z axis. The polar diagram

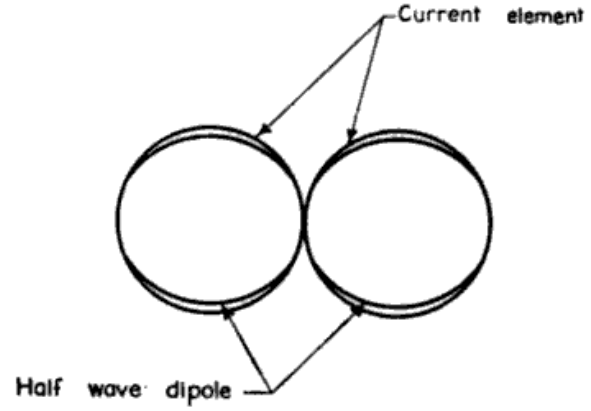


Figure 147.—Polar diagrams of a short current element and a half wave dipole in a plane containing them.

of a half wave dipole is plotted in Fig. 147 and may be compared with that of a short current element.

$$R_r = 73{\cdot}3 \text{ ohms} \tag{22}$$

where R_r is referred to the centre of the dipole

$$G_s = 1{\cdot}64 \tag{23}$$

$$P_{R_{max}} = 3{\cdot}46 \times 10^{-4}\,\lambda^2\,E^2 \tag{24}$$

$$A_{e_{max}} = 0{\cdot}13\lambda^2. \tag{25}$$

If a matched load is inserted at the centre of the dipole, then the voltage induced across it by an incident field E is given by

$$V = \frac{\lambda E}{2\pi}. \tag{26}$$

Effect of Finite Thickness of a Half Wave Dipole. The main effect of finite thickness is to increase the bandwidth and to decrease the resonant length (see Fig. 152). Thus fat dipoles must be used when wide bandwidth is required. The effect on S, R_r, G_s, etc. is of the second order.

The Dipole as a Circuit Element. The assumption of a sinusoidal current distribution leads to useful results in the case of the thin half wave dipole, but it gives no indication of the properties of an actual dipole. In some other cases the treatment is useless. For instance in the case of an aerial one wavelength long and fed at the centre, the current at the centre becomes zero and the impedance infinite.

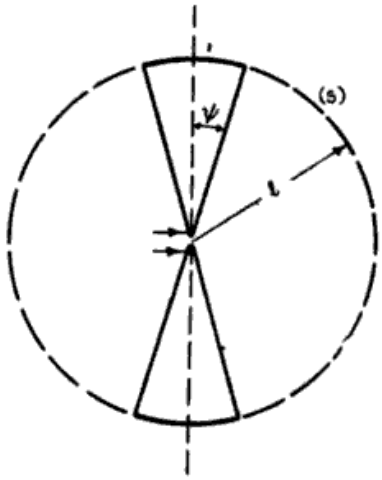


Figure 148.—Conical radiator with spherical ends and boundary sphere (S).

A method of overcoming this difficulty is used by Schelkunoff.[2] He considers a conical dipole of section given in Fig. 148, and assumes that the aerial and the surrounding

[2] S. A. Schelkunoff, *Electromagnetic Waves*, Van Nostrand, New York, 1943, pp. 441-79.

For a half wave conical dipole of half angle ψ

$$Q \doteqdot 2{\cdot}76 \log_{10} \frac{2}{\psi}. \tag{34}$$

The bandwidth is found from equation (7) to be

$$\frac{\triangle f}{f} = \frac{17}{K} \tag{35}$$

when the dipole matches the line at the resonant frequency.

Fig. 152 gives K, $\frac{\triangle f}{f}$, Q and percentage deviation p of the resonant length from $\lambda/2$ for a cylindrical dipole.

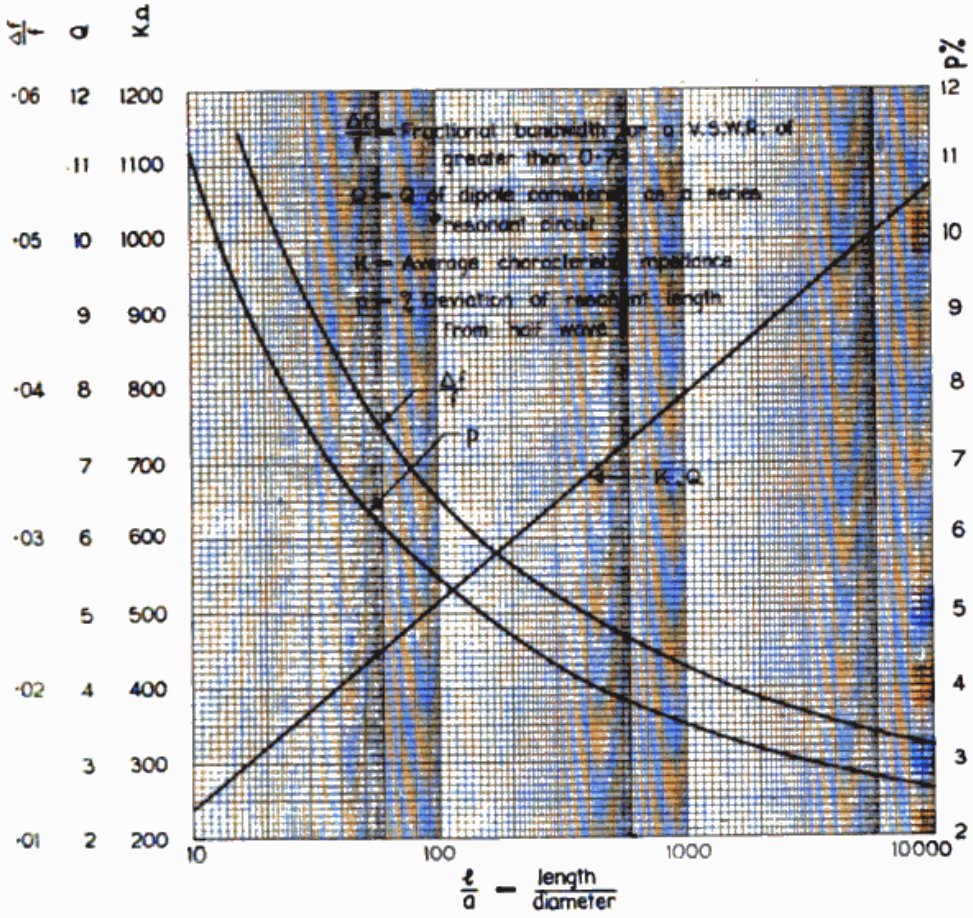


Figure 152.—Characteristic impedance, Q, and percentage deviation of resonant length from half wave, of a cylindrical dipole of length $2l$ and diameter $2a$ at half-wave resonance.

$$\left[2l = \frac{\lambda}{2}\left(1 - \frac{p}{100}\right)\right]$$

Methods of Feeding Dipoles

A dipole may be energised through a transmission line connected to its centre, and the feed is termed either resonant or non-resonant depending on whether or not standing waves exist on the line. It is in general desirable with high power radar transmitters to use a non-resonant feed since, for the same power, the maximum voltages appearing on the feeder are larger in the case of a resonant system.

Twin Line Feed. (*a*) *Resonant.* This consists of a twin line, an integral number of half wavelengths long connected between generator and dipole (Fig. 153). The impedance presented to the generator is thus the dipole impedance. The standing wave ratio on this feed depends on the ratio of characteristic

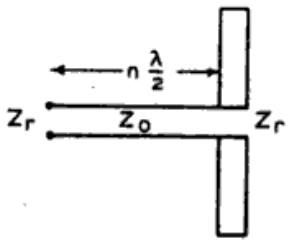


Figure 153.—Dipole with resonant feed consisting of a twin line an integral number of half-waves in length.

impedance Z_o of the line to the dipole impedance. Such feeds find their major application as connecting links between the elements of arrays.

(*b*) *Non-resonant.* If no standing waves are to exist on the line it must be matched to the dipole, for instance by making Z_o equal to the resonant dipole impedance. This, however, is in general difficult except when low powers are used and a line of very small spacing can be employed. If $Z_o \neq Z_r$, then a matching network such as a quarter wave transformer (Fig. 154) or a system of stubs must be used. Methods of matching transmission lines to loads are described in Chapter VII.

be excited by spacing them along a waveguide. The coupling to the guide and therefore the amount of energy abstracted from the guide can be adjusted by means of the probe.

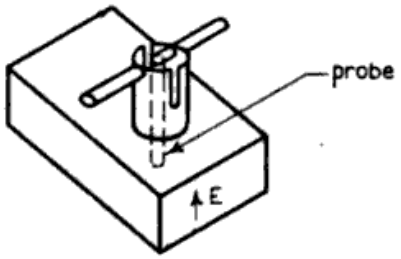


Figure 157.—Half-wave dipole excited from a waveguide by means of a probe situated parallel to the electric field.

Slot Radiators[3]

The concept of a slot in a conducting sheet as an efficient radiator is rather novel. It may be shown that a narrow slot approximately one half wavelength long in an infinite conducting sheet and energised at its centre AA', (Fig. 158)

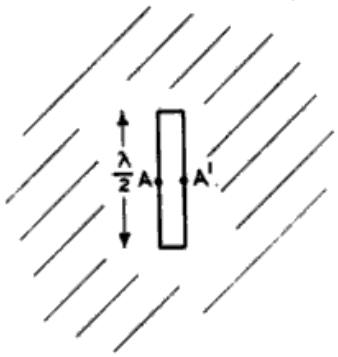


Figure 158.—Half-wave resonant rectangular slot in a conducting sheet.

behaves as an efficient resonant radiator of energy. To a first approximation it behaves in a similar fashion to a twin transmission line shorted at each end and fed at the centre, thus making it a resonant element. However, owing to the fact that the currents are not localised but spread out over the

[3] H. G. Booker, "Slot aerials and their relation to complementary wire aerials (Babinet's principle)," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 620-6.

sheet, the radiation from the slot is much greater than that from a two wire transmission line.

The properties of a slot may be determined from the equivalent properties of the complementary dipole by application of a modification of Babinet's principle which arises in optics.[3] Thus the field of a slot is obtained from that of the complementary dipole by interchanging electric and magnetic fields and reversing their direction on one side of the slot. That is, a slot is analogous to a hypothetical dipole consisting of a "perfect conductor of magnetic current" fed at its centre by a "magnetic current generator." A vertical slot has the same polar diagram and gain as a vertical dipole but produces a horizontally polarised field.

The radiation resistance of a narrow resonant slot is given by

$$R_{rd} \cdot R_{rs} = \frac{\eta^2}{4} \tag{36}$$

R_{rs} = slot resistance

R_{rd} = dipole resistance = 73 ohms.

Hence R_{rs} = 485 ohms. (37)

The bandwidth of a rectangular resonant slot is equal to that of the complementary dipole and is about one half that of a cylindrical dipole of diameter equal to the slot width. Since the resonant length of a rectangular slot is slightly less than $\lambda/2$ and obeys a similar law to that of a half wave dipole (see Fig. 152) we have for a slot of length $2l$ and width b

$$Q \doteqdot 5{\cdot}5 \log_{10} \frac{\lambda}{b} - 2{\cdot}4 \tag{38}$$

$$\frac{\Delta f}{f} \doteqdot \frac{0{\cdot}03}{\log_{10} \frac{\lambda}{b} - 0{\cdot}43}. \tag{39}$$

The current induced in a matched load at the centre of a slot by a field E is given in terms of the voltage induced across a matched load connected to the complementary dipole by

$$I_s = \frac{2}{\eta} V_d. \tag{40}$$

$$\text{Similarly } I_d = \frac{2}{\eta} V_s \tag{41}$$

GH intercepts shunt currents and so presents a shunt load lumped at the centre of the slot. *GH* is a shunt slot with maximum coupling to the guide. This coupling may be reduced in either of the ways shown in Fig. 164.

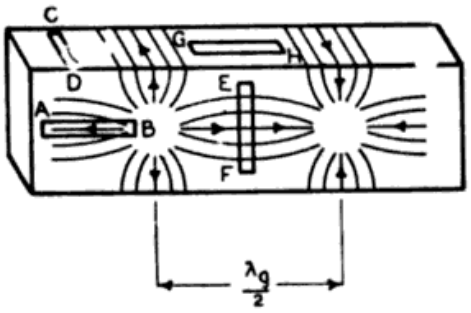


Figure 163.—Resonant slots in a waveguide. *EF* and *GH* are series and shunt slots respectively, while *AB* and *CD* are non-radiating slots.

The conductances presented to the guide by the slots of Fig. 164 (*a*) and (*b*) expressed as a ratio to the "characteristic conductance" of the guide are[4]

$$g = k \sin^2 \left(\frac{\pi x}{a}\right) \tag{44}$$

$$\text{and } g = k \sin^2 \phi \tag{45}$$

respectively, where

$$k = 2{\cdot}09 \frac{\lambda_g}{\lambda} \cdot \frac{a}{b} \cos^2 \left(\frac{\pi}{2} \frac{\lambda}{\lambda_g}\right) \tag{46}$$

λ and λ_g are the free space and guide wavelengths respectively.

Figure 164.—(*a*) and (*b*) show how the coupling of a shunt slot to a waveguide may be varied by either displacing it towards the centre of the broad face of the guide or by suitably orienting it in the narrow face.

If $g = 1$ and the guide is shorted one quarter wavelength from the end of the slot, all the energy supplied to the guide is

[4] W. H. Watson, "Resonant slots," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 747-77.

Dielectric Rod Aerials

The use of a dielectric rod as a leaky waveguide radiator is of relatively recent origin and, although it is not an array in the sense of an aggregate of elementary radiators, it may be considered as the limiting case of an end-fire array and so has been included in this section.

If in (51) we let l tend to zero and keep a (i.e. nl) constant, we obtain a line distribution of uniform amplitude and with a progressive phase lag of 360 degrees per wavelength along the array. This gives

$$S(\theta) = \frac{\sin\left(\frac{2\pi a}{\lambda}\sin^2\frac{\theta}{2}\right)}{\frac{2\pi a}{\lambda}\sin^2\frac{\theta}{2}} \tag{57}$$

$$G_s = \frac{4a}{\lambda} \quad (a \gg \lambda). \tag{58}$$

It is interesting to note that this is the same as that of a long end-fire array of point sources, $\lambda/4$ apart.

The dielectric aerial provides a method of realising the required phase distribution, although in general it will not give a uniform amplitude distribution. It consists of a tapered dielectric waveguide (Fig. 167) propagating a transverse electric wave and adjusted so that the velocity of the wave is equal to the free space velocity by shaping the rod, balancing the dielectric retardation against the waveguide acceleration.

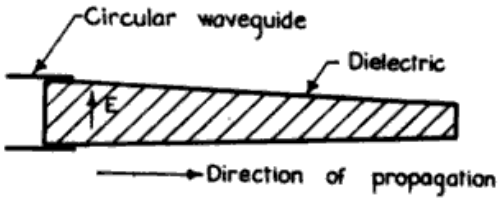


Figure 167.—Dielectric rod radiator.

Leakage from the rod, which may be fed from the wide end by matching it to a circular waveguide, causes it to act as an end-fire array. The power gain obtainable is slightly less

array. For a continuous, in phase, line distribution of length a

$$S(\theta) = \frac{\sin\left(\frac{\pi a}{\lambda}\cos\theta\right)}{\frac{\pi a}{\lambda}\cos\theta}, \tag{62}$$

$$G_s = \frac{2a}{\lambda}. \tag{63}$$

Comparing this with (61), it is seen that a broadside array of elements $\lambda/2$ apart has the same gain as a continuous line distribution of the same length.

Beam Width. For large n, the width of the main lobe between zeros is

$$\Delta = \frac{2\lambda}{nl} \text{ radians} \tag{64}$$

$$\doteqdot \frac{120\lambda}{a} \text{ degrees} \tag{65}$$

where a = overall length. The total beam width to half field points on the polar diagram is given approximately by

$$\delta \doteqdot \frac{70\lambda}{a} \text{ degrees.} \tag{66}$$

Rectangular Arrays

The most general array is a three dimensional lattice of elements, and the characteristics of such arrays may be determined from the following rule.

Array of Arrays Rule. An aerial array may be considered as a combination of elementary arrays forming an array of arrays. The directional characteristics of such an array is given by the directional characteristics of the elementary array multiplied by the directional characteristics of the combination of ele-

being 180 degrees out of phase with the array. The spacing of the elements from the reflector is not critical, but for very close spacing the impedance of the elements will be very low. In order to satisfy (49) the spacing should be $\lambda/4$ or less.

In Fig. 170 is plotted the radiation resistance of a $\lambda/2$ dipole situated at a distance d from and parallel to an infinite reflecting sheet together with the directivity, D, of the system

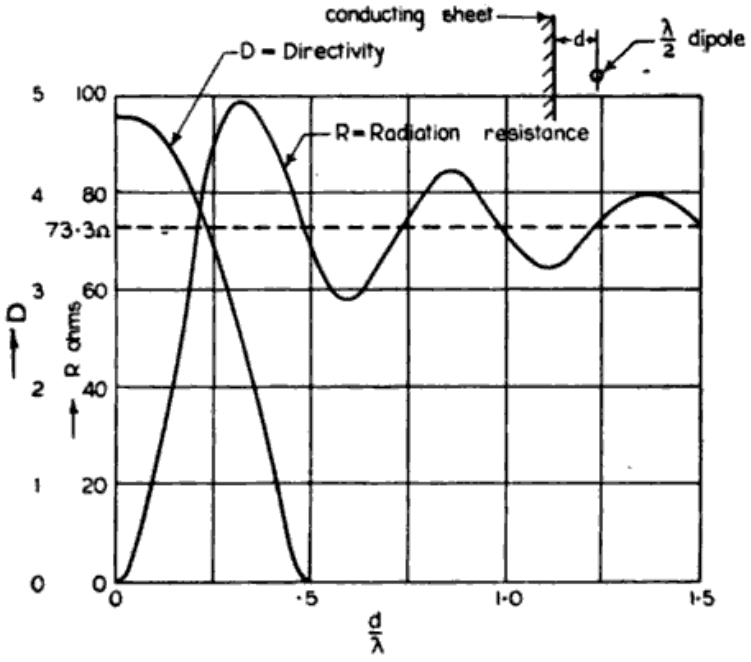


Figure 170.—Radiation resistance and directivity of a half-way dipole at a distance d from, and parallel to, an infinite conducting sheet.

relative to that of an isolated dipole. The curve for gain is substantially the same as that for directivity except at small spacings, when it becomes less due to ohmic resistance. The equation to the directivity curve is

$$D = \frac{R_\infty}{R}\left(2 \sin \frac{2\pi d}{\lambda}\right)^2 \qquad (70)$$

where $R_\infty = 73{\cdot}3\omega$ = radiation resistance of an isolated dipole, and R = radiation resistance of the dipole and reflector. If R equals the sum of the sum of the ohmic and radiation resistances presented at the dipole terminals, (70) gives the gain

Feeding of Broadside Arrays

The simplest method of feeding broadside arrays is by spacing the elements along a transmission line or waveguide. A typical case is illustrated in Fig. 171. Each alternate dipole is reversed to compensate for the 180 degrees change in phase along a $\lambda/2$ length of line. This is a resonant feed, since standing waves will exist on the feed line. An alternative to reversing the dipoles is to transpose the line between dipoles.

If the dipoles are spaced, say, every 200 degrees along the line instead of 180 degrees, the reflections back along the line from each element tend to cancel and the standing waves on

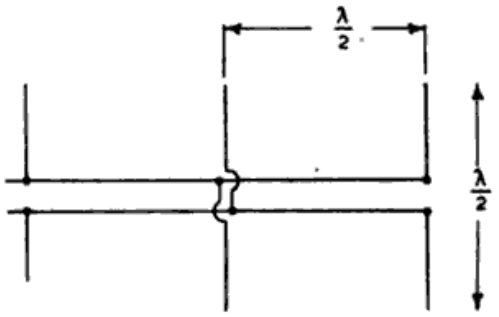


Figure 171.—Resonant-fed array of parallel half wave dipoles.

the line are reduced. The array is fed from one end, each element extracting a small fraction of the power, and the residue (generally about 10 per cent) is absorbed by a matched load at the other end. However since each element has to extract a progressively greater percentage of the power in order to maintain a given excitation, they must present a progressively smaller impedance to the line. This is generally difficult to arrange. Also the presence of the elements modifies the phase along the line and so the phase difference between the elements will vary from the desired amount. This effect however is usually small. A slight disadvantage is that if the elements are spaced every 200 degrees along the line (the line being transposed between radiators) there is a progressive phase difference of 20 degrees between radiators which will

being obtained by placing alternate slots on opposite sides of the centre line. Coupling is varied by adjusting the distance of the slot from the centre line [see equation (44)]. Slots may be excited by any of the methods described previously.

For non-resonant feeds the elements could be spaced, say, 200 degrees (in the guide) and the array terminated in a matched load. In this case, again, the beam will be deflected from the normal to the array. Because of the ease of varying the coupling of individual radiators, a non-resonant array is more easily obtainable in waveguide than conventional transmission line. A method of calculating the required element impedances is given in a report by Kaiser.[8]

If in Fig. 172, $l = \lambda_g/2$, (where λ_g = guide wavelength), and the guide is shorted $\lambda/4$ beyond the centre of the end radiator, a resonant system analogous to that of Fig. 171 is obtained, the only difference being that the element spacing is greater than one half wavelength in free space.

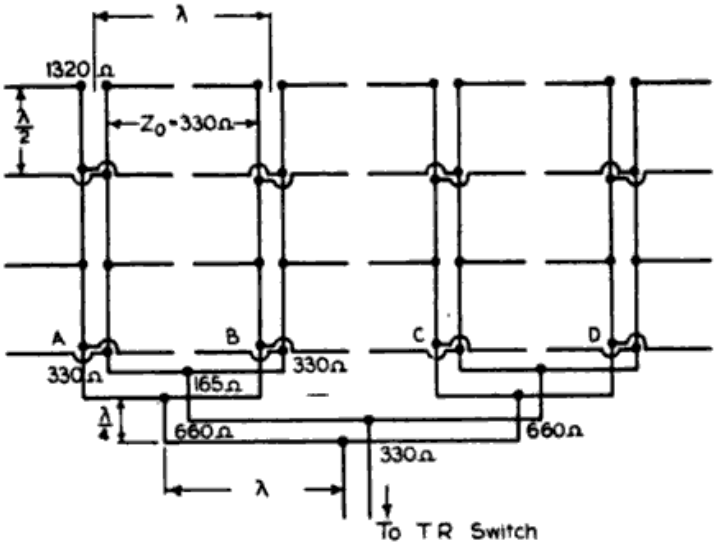


Figure 173.—Thirty-two element array of half-wave end-fed dipoles used with radar-set type LW/AW Mk.I showing the feeder system which uses 330 ohm twin transmission line throughout.

An example of the feeding of a typical rectangular array is given in Fig. 173. This array was used with the Australian

[8] T. R. Kaiser, "The design of a waveguide-fed slot array to produce a cosec² θ polar diagram," C.S.I.R. Radiophysics Laboratory, Report TI 223, January, 1946; A. L. Cullen and F. K. Goward.[7]

impedance of the element and the mutual impedances between it and all other elements. Tables of mutual impedances of $\lambda/2$ dipoles for various spacings and displacements are given by Terman,[6] (p. 779) and a more extensive table by Telecommunications Research Establishment.[9] If the number of elements is large, mutual effects tend to cancel and may be neglected.

Arrays of Dielectric Rods

The gain of a rectangular array of elements backed by a reflector is seen by (72) to be proportional to the area of the array. A somewhat higher gain may be obtained if each element of the array is itself an end-fire array. This implies that the absorption cross section of the array is greater than the actual area. The increase is greatest for small arrays. An example of this type is a broadside array of dielectric rods. Such a system has been developed by Bell Telephone Laboratories.

Yagi Arrays

The Yagi is a particular type of end-fire array consisting of a single fed element and a number of parasite elements. Currents are induced in the latter and the element spacings and parasitic impedances are arranged so that these currents produce fields which reinforce along the array.

If a parasitic element causes maximum radiation along the line from the driven element to the parasitic it is termed a director, and if in the opposite direction, a reflector. A dipole of length slightly less than the resonant length (i.e. with a capacitive impedance) behaves as a director, while if it is slightly greater (inductive) it behaves as a reflector. The optimum values of the lengths of such elements depend on the spacing, the length being adjusted so that the reactive component of the impedance causes the parasitic current to lag or lead the current in the fed element by the desired amount.

[9] Telecommunications Research Establishment, Mathematics Group, "Mutual impedance of half wave aerials," T.R.E. Report M3, January, 1941; R. A. Smith, *Aerials for Metre and Decimetre Wavelengths*, Cambridge University Press, 1949, pp. 10-1.

of solving the diffraction problem is advanced. However, if the aperture is large with respect to the wavelength, the simple application of Huygens' principle gives results which agree closely with experiment.

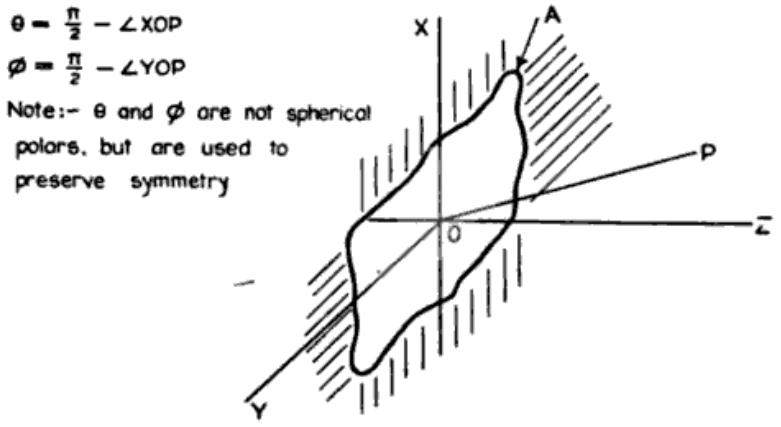


Figure 174.—Coordinate system for specifying the space factors (θ, ϕ) of an illuminated plane aperture in an opaque sheet situated in the XOY plane. The direction (θ, ϕ) is that of the vector OP which has direction cosines sin θ and sin ϕ relative to the X and Y axes.

The space factor for a plane aperture A (Fig. 174) may be shown to be

$$S(\theta,\phi) = \left| \int_A F(x,y)\, e^{j\frac{2\pi}{\lambda}(x\sin\theta + y\sin\phi)}\, dA \right| \qquad (74)$$

where $F(x,y) = f(x,y)e^{jg(x,y)}$ defines the aperture distribution in both amplitude and phase. The integral of (74) gives the amplitude and phase in the diffracted field and its amplitude gives $S(\theta,\phi)$. The space factor of (74) does not quite satisfy the definition of section 1 in that its maximum value, in general, is not unity. If the aerial produces a sufficiently narrow beam (i.e. if A is large) then we may write

$$\sin\theta \doteqdot \theta$$

$$\sin\phi \doteqdot \phi$$

and (74) becomes

$$S(\theta,\phi) = \left| \int_A F(x,y)\, e^{j\frac{2\pi}{\lambda}(x\theta + y\phi)}\, dA \right|. \qquad (75)$$

Rectangular Aperture

Consider a rectangular aperture of dimensions a and b in the x and y directions respectively, illuminated so that the aperture distribution may be expressed by

$$F(x,y) = F_1(x) \cdot F_2(y). \tag{80}$$

This gives

$$S(\theta,\phi) = \left| \int_{-a/2}^{a/2} F_1(x) e^{j\frac{2\pi}{\lambda} x\sin\theta} dx \right| \left| \int_{-b/2}^{b/2} F_2(y) e^{j\frac{2\pi}{\lambda} y\sin\phi} dy \right| \tag{81}$$

$$= S(\theta) \times S(\phi),$$

where

$$S(\theta) = \left| \int_{-a/2}^{a/2} F_1(x) e^{j\frac{2\pi}{\lambda} x\sin\theta} dx \right| \tag{82}$$

and

$$S(\phi) = \left| \int_{-b/2}^{b/2} F_2(y) e^{j\frac{2\pi}{\lambda} y\sin\phi} dy \right| \tag{83}$$

are the one dimensional space factors in the planes $\phi = 0$ and $\theta = 0$ respectively. Thus we need only consider the one dimensional polar diagrams.

Similarly we may write

$$\gamma = \gamma_1 \cdot \gamma_2 \tag{84}$$

where γ_1 and γ_2 are the vertical and horizontal one dimensional gain factors respectively.

$$\gamma_1 = \frac{\left| \int_{-a/2}^{a/2} F_1(x) dx \right|^2}{a \int_{-a/2}^{a/2} \left| F_1(x) \right|^2 dx}. \tag{85}$$

Graphs of $S(\theta)$ and gain factors for a number of distributions $F(x)$ are given by Ryan.[11]

Uniformly Illuminated Rectangular Aperture

From (86) we obtain for the space factor in the plane $\phi = 0$ (Fig. 174)

$$S(\theta) = \frac{\sin\left(\frac{\pi a}{\lambda}\sin\theta\right)}{\frac{\pi a}{\lambda}\sin\theta} \tag{89}$$

$$\doteqdot \frac{\sin\frac{\pi a\theta}{\lambda}}{\frac{\pi a\theta}{\lambda}}$$

which is the same as that for the continuous, in-phase, linear array of p. 246.

From equation (76)

$$G_s = \frac{4\pi ab}{\lambda^2},$$

where a and b are the sides of the aperture.

Equation (89) gives the beam widths in the plane $\phi = 0$ to be

$\Delta \doteqdot 120\frac{\lambda}{a}$ degrees between the first two minima,

$\delta \doteqdot 70\frac{\lambda}{a}$ degrees to the half field points,

amplitude of first side lobe = 21 per cent.

The important results are that gain is proportional to area and that beam width is inversely proportional to linear dimensions.

Circular Aperture With Uniform Illumination

It may be shown that

$$S(\theta) = \frac{J_1\left(\frac{\pi a}{\lambda}\sin\theta\right)}{\frac{\pi a}{\lambda}\sin\theta}. \tag{90}$$

Also $G_s = \frac{\pi^2 a^2}{\lambda^2}$ from equation (76).

The case of a circular aperture having a circularly symmetrical amplitude distribution given by

$$f(r) = 1 - \left(\frac{4r^2}{a^2}\right)^t,$$

where r is distance from the centre of the aperture, is evaluated by Ryan[11] for $0 < t < 2$.

Parabolic Reflectors

The use of a parabolic reflector to obtain an equiphase distribution across a plane aperture is very common. Such reflectors may be either paraboloids of revolution with circular apertures or parabolic cylinders with rectangular apertures. The former may be cut so that the projected aperture is rectangular in shape. The source of excitation or feed in the case of the paraboloid must be an effective point source and in the case of a parabolic cylinder it must act as an effective line source.

If the field distribution in the aperture of the parabolic reflector is known by calculation or measurement, the radiation characteristics may be calculated by the principles and results discussed in the previous sections.

Paraboloids. For the theoretical behaviour of paraboloids with various feeds the reader is referred to the papers listed below.[13]

The condition of uniform phase across the aperture of a paraboloid is automatically obtained by using a small (effectively point) source at its focus ; however, a uniform amplitude distribution is never obtained owing to the peculiar polar diagram which would be required from the source. If the focus were in the plane of the aperture, four times more energy would have to be directed at the edge than at the centre.

[13] C. C. Cutler, "Parabolic antenna design for microwaves," *Proc. I.R.E.* Vol. 35, 1947, pp. 1284-94; E. G. Brewitt-Taylor, "A detailed experimental study of the factors influencing the polar diagram of a dipole in a parabolic mirror," *J. Instn. Elect. Engrs.*, Vol. 93, Part IIIA, 1946, pp. 679-82; H. T. Friis and W. D. Lewis, "Radar antennas," *Bell Syst. Tech. J.*, Vol. 26, 1947, pp. 219-317; R. Dabord, "*Reflecteurs et lignes de transmission pour ondes ultra-courtes*," *Onde Elect.*, Vol. 11, 1932, pp. 53-82.

The effect of the tapered distribution is to lower the gain, at the same time widening the beam and reducing side lobes. By using a suitable aperture distribution side lobes may be entirely eliminated at the expense of a broader beam. Fig. 177

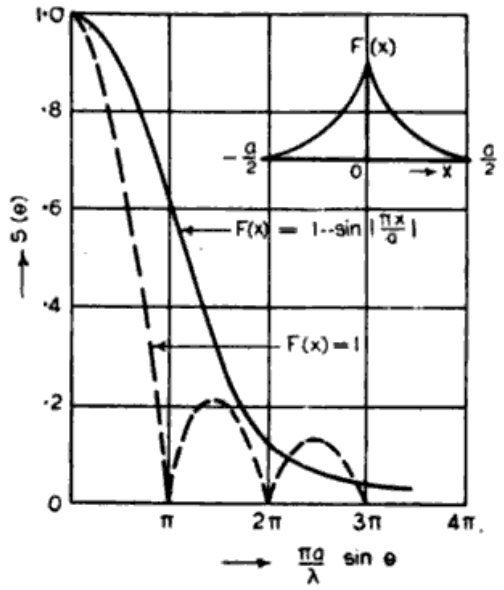


Figure 177.—Illustrates how a suitably tapered aperture illumination produces a diagram with no side lobes.

illustrates such a case, where the distribution is given by $F(x) = 1 - \sin\left|\frac{\pi x}{a}\right|$. The diagram corresponding to $F(x) = 1$ is given also for comparison.

The simplest feed is a single dipole and it is shown theoretically by Darbord[13] that the optimum dish is one with its focus

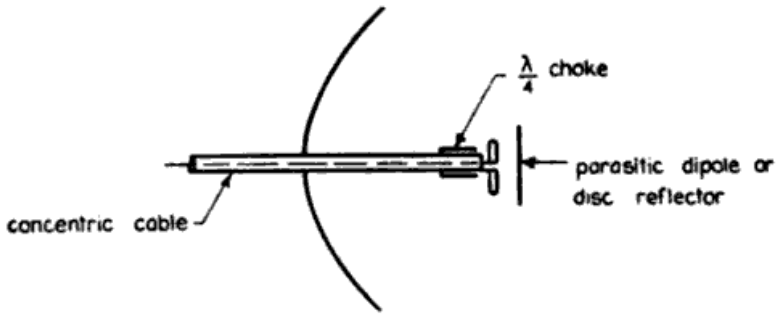


Figure 178.—Paraboloid with dipole feed.

in the aperture plane. However, owing to the small degree of illumination at the edges, it may be economical to have the focus a short distance in front. The gain factor of such a

The ends of the parabolic cylinder may be closed by parallel plates without materially affecting the performance if the electric vector is parallel to the axis of cylinder. If however the electric vector is parallel to the end plates, waveguide propagation may take place within the cylinder and allowance may have to be made for this, especially in short cylinders.

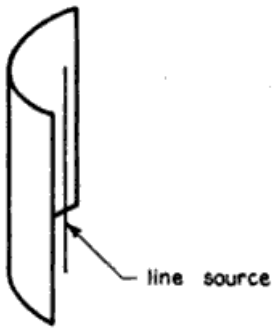


Figure 181.—Parabolic cylinder fed by a line source.

Cheese Aerial

This is the name given to a parabolic cylinder closed at its ends by parallel plates and fed with the electric vector parallel to the plates. Fig. 182 shows a cheese fed by a sectoral horn (see p. 271) excited in the TE_{10} mode.

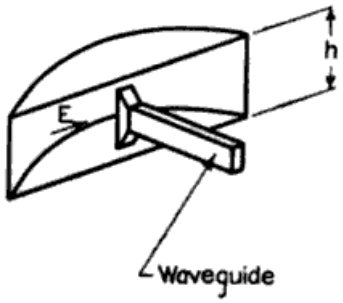


Figure 182.—Cheese aerial with a horn feed.

Unless the cheese is sufficiently narrow many modes of propagation may occur within it; however, by using a horn feed of aperture equal to the cheese height, the lowest TE mode only is propagated and the amplitude distribution across the short dimension of the aperture follows a cosine law corres-

at the narrow dimension of the horn, which causes a very wide beam having its phase centre almost in the aperture. This distribution will of course be tapered. The effective phase centre for illumination of the narrow dimension of the cheese is near the apex of the horn. If the height of the cheese is small, it may be fed by an open ended waveguide.

Pill Box Aerial[14]

If a parabolic cylinder closed at both ends is fed by a wave having its electric field parallel to the cylinder, only one wave, the *TEM*, can be propagated, and the phase velocity between the plates equals the free space velocity. A typical pill box with waveguide feed is shown in Fig. 185. The distribution across the narrow dimension in this case is constant.

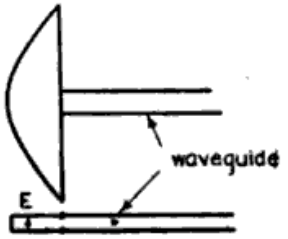


Figure 185.—Pill-box aerial with waveguide feed.

Both the pill box and the narrow cheese provide a good approximation to a line source with phase centre just inside the aperture, producing a very narrow beam in their own plane and a wide beam in the plane at right angles.

Bandwidth of Reflector Type Aerials

Since the feed can be made with a relatively large bandwidth, the bandwidth of a reflector type aerial is considerably greater than that of an array. It is however slightly less than the bandwidth of the feed since some of the reflected energy is picked up by the feed and causes standing waves on the feeder. This standing wave can be eliminated at one frequency by a matching adjustment on the feeder, but will be present for other frequencies owing to the change in phase of

[14] H. T. Friis and W. D. Lewis, "Radar antennas," *Bell Syst. Tech. J.*, Vol. 26, 1947, pp. 219-317.

the horn angle, at some point the gain will commence to decrease. In the case of the TE_{10} wave this occurs when the phase difference between the fields at the edge and centre of the aperture is 90 degrees.

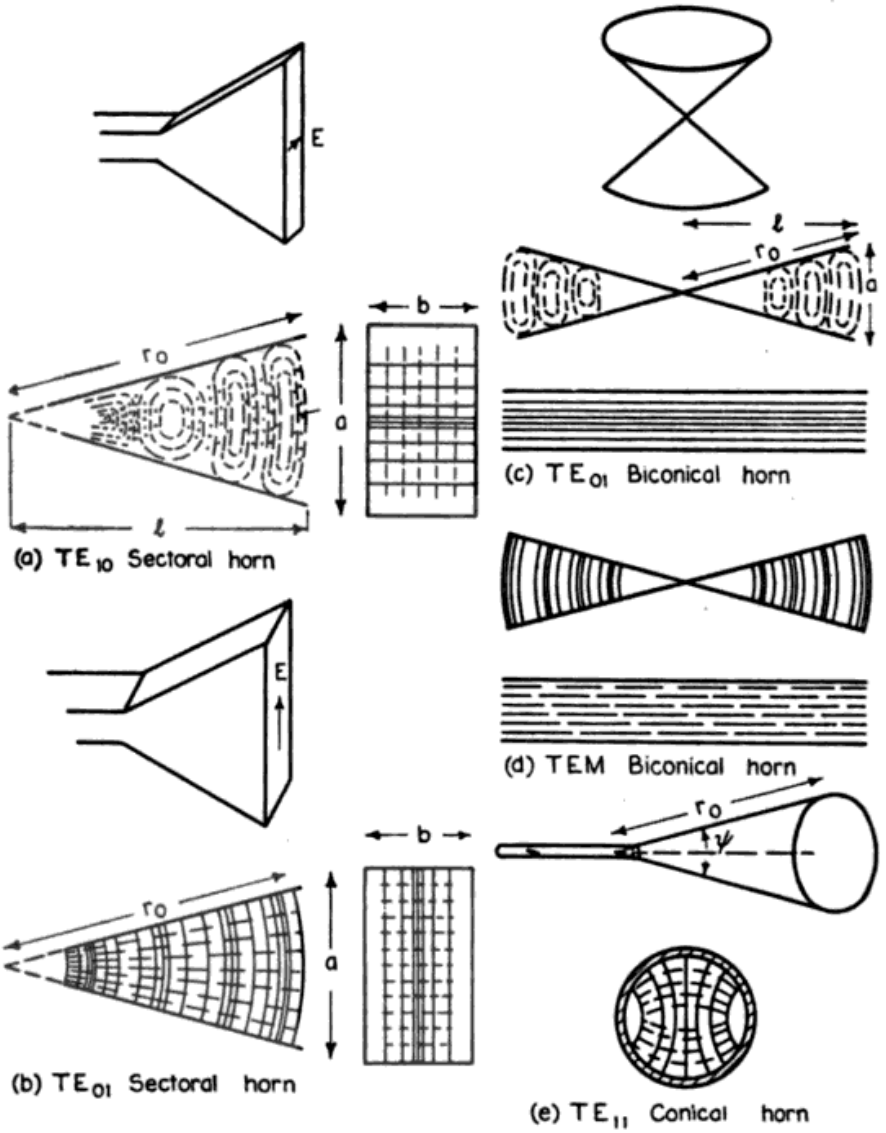


Figure 188.—Field configurations in sectoral, conical, and biconical electromagnetic horns.

Sectoral Horn Excited by the TE_{01} Mode [Fig. 188 (*b*)]. The assumption that the aperture distribution is the same as that which would exist at that point in an infinite horn of the same

Conical Horn.[19] For a TE_{11} conical horn [Fig. 188 (*e*)] fed by a circular waveguide, the optimum dimensions are given by

$$r_o = \frac{0{\cdot}3\lambda \cos \frac{1}{2}\psi}{1 - \cos \frac{1}{2}\psi}. \tag{100}$$

Biconical Horn.[20] Fig. 188 (*c*) and (*d*) illustrate the form of the electromagnetic field in a biconical horn for the TE_{01} and the TEM mode respectively. It will be noted that the former corresponds to the TE_{10} mode in a sectoral horn, and the latter to the TE_{01} mode; thus the optimum dimensions are the same in both cases.

If we consider a meridian plane only, the gain factor corresponding to the aperture distribution is the same as for the corresponding sectoral horn and may be found from equations (93) and (96) or from Fig. 189 for given horn dimensions. The gain is then given by[21]

$$G_s = \gamma_a \cdot \frac{2a}{\lambda} \tag{101}$$

since $2a/\lambda$ is the gain corresponding to uniform phase and amplitude aperture distribution [see equation (63)].

For optimum horns,

TEM	TE_{01}	
$a = \sqrt{2\lambda r_0}$.	$a = 1{\cdot}78\sqrt{\lambda r_0}$.	(102)
$G_s = 1{\cdot}6\frac{a}{\lambda}$.	$G_s = 1{\cdot}26\frac{a}{\lambda}$.	(103)

Bandwidth of Horns. Effectively the horn may be regarded as a tapered matching section between the guide and free space, thus if the taper is not too rapid the impedance of the horn will not vary rapidly with change in frequency.

[19] G. C. Southworth and A. P. King, "Metal horns as directive receivers of ultra short waves," *Proc. I.R.E.*, Vol. 27, February, 1939, p. 95.

[20] W. L. Barrow, L. J. Chu and J. J. Jansen, "Biconical electromagnetic horns," *Proc. I.R.E.*, Vol. 27, December, 1939, p. 769.

[21] Barrow, Chu and Jansen have plotted gain relative to a current element against flare angle for various values of r_o/λ; however, the figures for gain are 4·8 times higher than expected from (101) i.e. some 3 or 4 times higher than that of an in-phase, uniform amplitude, line source of the same aperture.