

## CHAPTER 26 - ACOUSTIC MEASUREMENT

Contents	Page
Introduction .....	1
Sound Intensity .....	2
Sound Pressure Level .....	2

### Figure

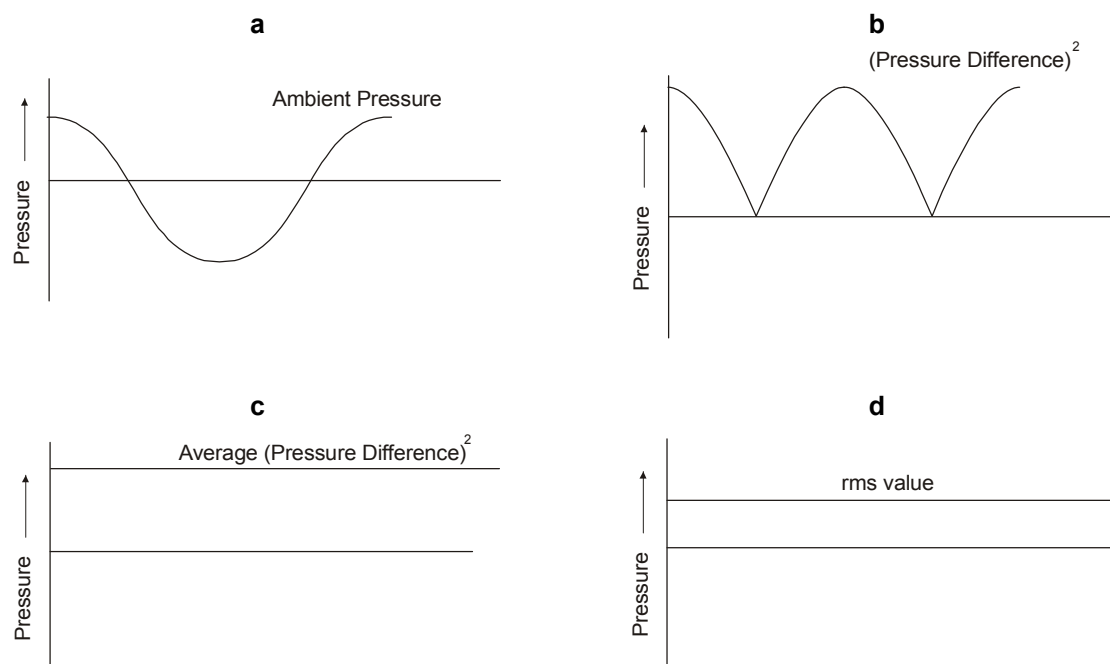
13-26 Fig 1 Root-mean-square Derivation Sound Intensity .....	1
---------------------------------------------------------------	---

### Introduction

1. Sound consists of small variations in pressure above and below the ambient pressure with respect to time. It may be depicted diagrammatically as a sine curve reflecting the gradual increase in pressure to a maximum and its subsequent decrease to a minimum. The maximum excursion of pressure difference from the ambient level is called the amplitude, but for practical purposes some kind of average value over time is required that gives equal weighting to both the rarefaction and compression phases. This 'average' value is known as the root-mean-square (RMS) value, and its derivation is described below.

2. **Root-mean-square Value.** Fig 1a shows a graph of the sinusoidal sound pattern with amplitude on the y-axis. The first stage, (Fig 1b), is to square all the values of amplitude which has the effect of making the negative values positive. The squared values are then averaged, (Fig 1c), to produce a level value. Finally the square root of the average is taken (Fig 1d). As with all other pressures, the units are newtons per square metre ( $\text{Nm}^{-2}$ ) or Pascals (Pa).

**13-26 Fig 1 Root-mean-square Derivation Sound Intensity**



### Sound Intensity

3. Sound intensity is a measure of the power transmitted per unit area, the area being at right angles to the direction in which the sound is propagating. For unhindered sound, away from the source, the intensity is proportional to the square of the pressure i.e.:

$$I = kp^2$$

where k is a constant determined by the medium through which the sound is travelling (e.g. for air at atmospheric pressure and 20 °C,  $k = 1/410$ ). The term 'sound intensity' is used in some reference books, but it is not used as a measure in underwater acoustics.

### Sound Pressure Level

4. In underwater acoustics, measurements are always in terms of sound pressure level (SPL). Sound pressure levels cover a very wide range of values, for example the range from the threshold of hearing to the onset of pain is from  $2 \times 10^{-5}$  Pa to 20 Pa. Using a linear scale over this range would be cumbersome, and so a logarithmic (decibel) scale is used relative to a datum pressure level. The datum pressure that is chosen for sound in air is  $2 \times 10^{-5}$  Pa, which is equivalent to the lowest sound pressure at 1,000 Hz, detectable on average by people with normal hearing. In underwater measurement, the datum pressure level is the micropascal, ( $1 \times 10^{-6}$  Pa).

5. As the energy contained in a wave is proportional to the amplitude squared, and if p is the sound pressure to be measured, then if the transmission was to take place in water:

$$\begin{aligned} \text{SPL} &= 10 \log \frac{p^2}{(1 \times 10^{-6})^2} \text{ dB} \\ &= 20 \log \frac{p}{1 \times 10^{-6}} \text{ dB} \end{aligned}$$

6. Note that a significant change in sound pressure level is represented by only a small change in decibel notation. If the pressure from a noise source is doubled, this equates to  $10 \log 2$  in dB, ie 3 dB. Thus, if the noise level from, say, a jet aircraft with four engines running was 140 dB, the same aircraft running on two engines would generate a level of 137 dB. A one or two dB difference in the measurement of acoustic pressure may therefore be significant in target detection.

## CHAPTER 27 - OCEANOGRAPHY

Contents	Page
Introduction .....	1
GEOLOGY AND STRUCTURE.....	1
General .....	1
Atlantic Margin .....	2
Pacific Margin .....	3
SEAWATER .....	3
Physical Properties.....	3
Waves.....	4
Currents.....	5
The Biological Environment.....	6

### Table of Figures

13-27 Fig 1 Cross-section of the Earth's Crust between Africa and South America .....	1
13-27 Fig 2 Typical Topography of an Atlantic Style Ocean Margin.....	2
13-27 Fig 3 Depth/Temperature Profiles for Different Latitudes .....	4
13-27 Fig 4 Characteristics of an Ocean Wave .....	4
13-27 Fig 5 Generalized Surface Currents .....	5

### Table

Table 1 Sea State and Beaufort Wind Scale.....	7
------------------------------------------------	---

### Introduction

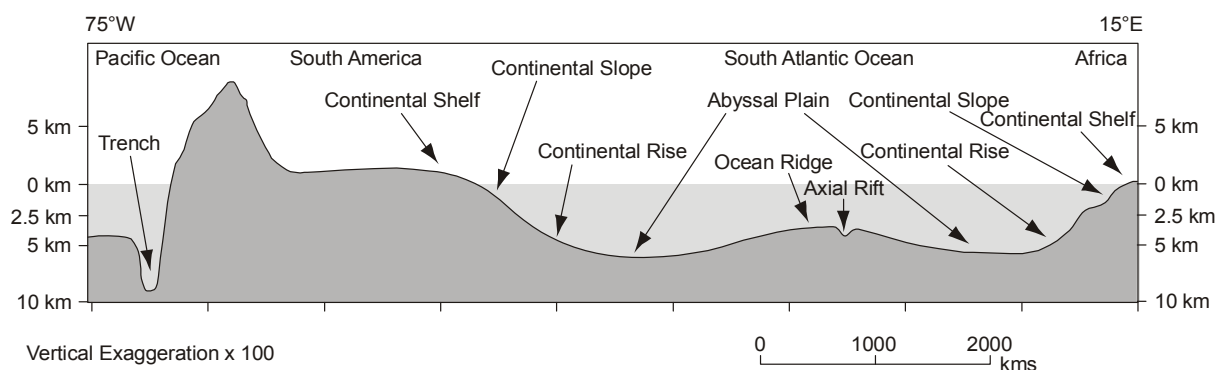
1. Maritime Patrol Aircraft are tasked with the location of submarines using acoustic sensor equipment. The aim of this chapter is to outline those aspects of the oceanic environment which have a bearing on the transmission of sound. Details of how acoustic systems are affected by this environment are covered in Volume 13, Chapter 28.

## GEOLOGY AND STRUCTURE

### General

2. The oceans are not simply depressions in the Earth's surface filled with water, rather they have a fairly complex structure. Fig 1 shows a cross section of the Earth's crust between Africa and South America and just into the eastern Pacific Ocean.

**13-27 Fig 1 Cross-section of the Earth's Crust between Africa and South America**



3. In the centre of the Atlantic Ocean is what amounts to a mountain chain - the Mid-Atlantic Ridge. This is part of a ridge system which extends from North of Iceland southwards into the South Atlantic, eastwards into the Indian Ocean, then northwards again in the East Pacific before disappearing under the North American continent. Its total length is over 50,000 km and its width between 1,000 and 4,000 km. The ridge generally rises to about 2 km, occasionally 5 km, above the flanking plain, and the crest is normally about 2½ km below sea level. The Mid-Atlantic Ridge has an axial rift and it is here that new ocean crust is created. This crust formation leads to the Atlantic widening at a rate of 2 to 4 cm/year.

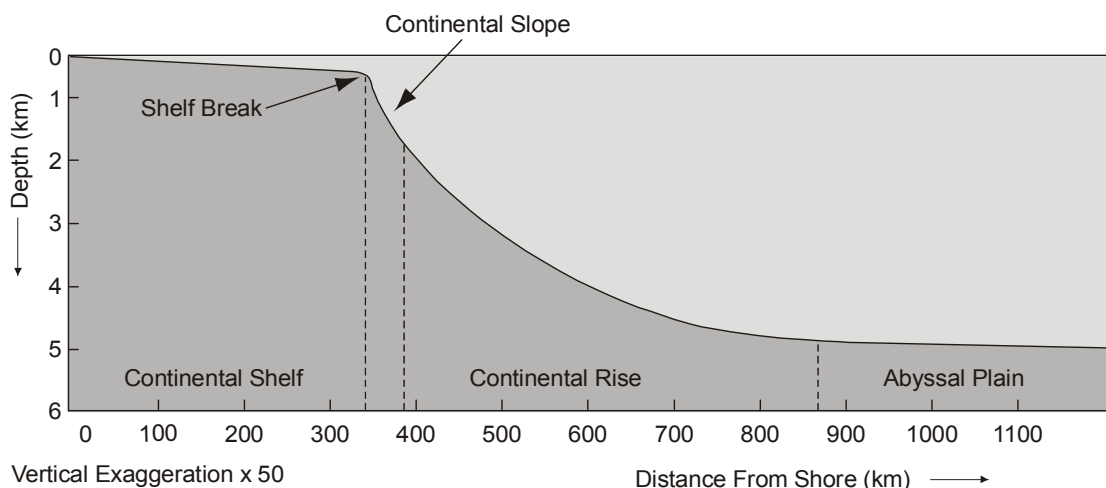
4. Adjacent to, and either side of, the ridge lie the abyssal plains which are in general relatively flat areas. This flatness is primarily due to a thick layer of sediment overlaying the rough topography which has been generated at the ridge. In places this flatness is punctuated by abyssal hills and seamounts. The majority of these are ocean floor volcanoes and most do not rise above sea level; where they do, they form oceanic islands. The Pacific Ocean is particularly well endowed with seamounts some of which are very large. Mauna Loa for example is about 100 km across at its base and rises to about 9 km above the plain, ie about as high as Mt. Everest. Many of these seamounts have flat tops and are known as guyots. It is believed that this flat-topped effect was caused by wave erosion at a time when the volcano was at or above sea level; with the passage of time, they have subsided below the surface.

5. The margins of the Atlantic and Pacific Oceans are rather different. The Pacific margin is characterized by seismic and volcanic activity whereas the Atlantic margin is aseismic. This is because the Pacific margin coincides with the edge of a crustal plate, whereas in the case of the Atlantic the ocean and the adjacent continents are on the same plate. This difference gives rise to characteristically different ocean floor topography.

### Atlantic Margin

6. Fig 2 shows the typical shape of an Atlantic style ocean margin. The continental shelf is geologically part of the continent rather than the ocean; the underlying rock type is not oceanic. The shelf is normally coated with sediments derived from the adjacent land. The width of the shelf can extend to as much as 1,500 km although for the most part, it is less than this, typically a few hundred kilometres. The surface is generally flat with an average gradient of only 0.1°. At the shelf break the water depth varies between 20 and 500 m with an average of around 130 m.

13-27 Fig 2 Typical Topography of an Atlantic Style Ocean Margin



7. **Continental Slope.** The continental slope marks a sharp increase in gradient to about  $4^\circ$  on average. The width of the slope is between 20 and 100 km and the base of the slope lies at a depth of between 1.5 and 3.5 km. Frequently the slope is cut by submarine canyons along which sediment is transported to the deep ocean. These canyons, which are somewhat similar to V-shaped river valleys on land, often start on the continental shelf and commonly coincide with the mouths of major rivers. The end of the continental slope marks the end of the continental crust and the beginning of the oceanic crust.

8. **Continental Rise.** The continental rise has a much gentler gradient than the slope, about  $1^\circ$ , and is built up of the sediments which have flowed down the slope. Typically the continental rise extends for about 500 km and reaches a depth of some 4 km before merging into the abyssal plain.

### Pacific Margin

9. The most significant difference between an Atlantic and a Pacific type margin is the presence of a very deep trench in the latter at the outer edge of the continental slope. In general the shelf is narrower in the Pacific, about 50 km. The shelf break tends to be more abrupt and the slope often has a steeper gradient, up to about  $10^\circ$  in places. The continental rise is missing and is replaced by a trench which marks the site where the oceanic crust is being subducted beneath the continental crust. It is this subduction which gives rise to seismic and volcanic activity. The trenches can reach depths of over 11 km, and a depth of 8 km would not be untypical. In places the trench may be partly filled with land derived sediments.

## SEAWATER

### Physical Properties

10. **Salinity.** Salinity is a measure of the dissolved solids in seawater. It is usually expressed in values of parts per thousand by weight (i.e.  $\text{gm kg}^{-1}$ ) and the symbol o/oo is used. Surface salinity represents a balance between an increase due to evaporation and freezing on the one hand, and a decrease due to precipitation, ice melting, and river influx on the other hand. Despite these effects, salinity values do not vary greatly; between 30 and 40 o/oo at the extremes. Minimum values occur in coastal regions with large river discharges and the maximum values occur in the Red Sea and Persian Gulf. In the open ocean surface salinity values are even more conservative, varying between about 33 and 37 o/oo. The maximum values tend to occur around the tropics, where the effect of evaporation is at its highest. In low and middle latitudes salinity decreases with depth in the first 600 to 1000 m, a zone known as a halocline, below which it becomes virtually constant at 34.5 to 35 o/oo.

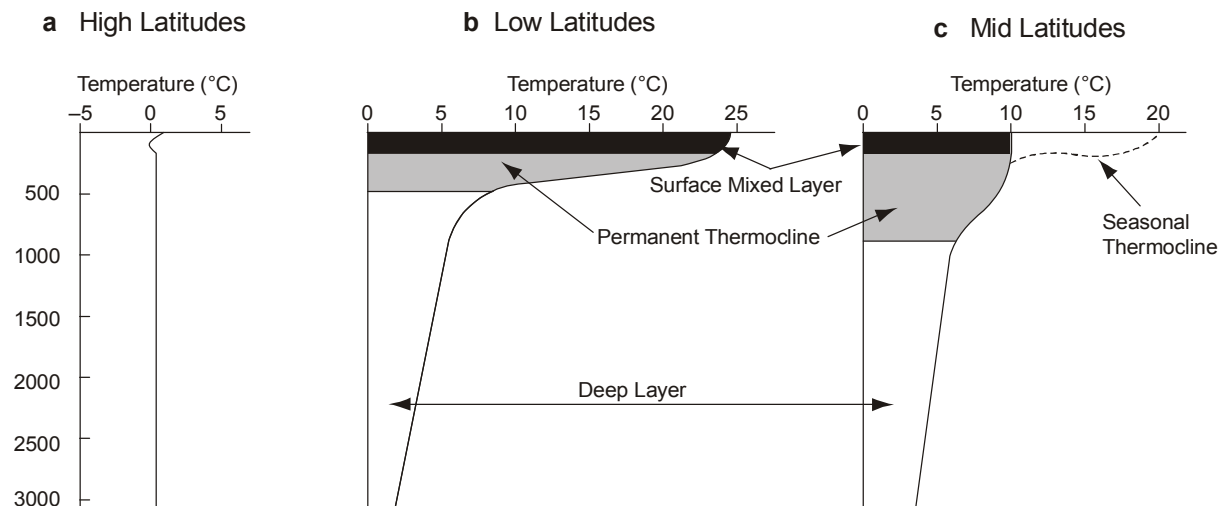
11. **Pressure.** Assuming constant density and a constant value for  $g$  (the acceleration due to gravity), then pressure varies virtually linearly with depth. For 99% of the oceans, density remains within  $\pm 2\%$  of its mean value, and variations in  $g$  are very much smaller than this so the linear relationship is a reasonable model. The change is approximately  $10^5 \text{ Nm}^{-2}$  (1 bar or 1 atmosphere) per 10 m (33 ft).

12. **Temperature.** The only agent for heating the oceans is incoming solar energy and the majority of this is absorbed within metres of the ocean surface. All of the infra-red radiation is absorbed within one metre and only about 2% of the incident energy reaches 100 m. A small amount of heat is transmitted to depth by conduction but mixing caused by wind and waves is the main mechanism for the transfer of heat. This turbulence generates a mixed surface-layer which may be up to 200 m thick depending upon surface conditions, and therefore upon the season.

13. **Permanent Thermocline.** Below this mixed layer, and down to about 1,000 m, the temperature falls rapidly. This region is known as the permanent thermocline. Underneath this, seasonal variations are virtually non-existent and there is a much shallower temperature gradient with temperature falling to between 0.5 °C and 1.5 °C. This 3-layer model of the ocean temperature structure is illustrated in Fig 3 showing the variations with latitude.

14. **Season and Latitude Effects.** In high latitudes (above 60° N) the permanent thermocline is missing. In mid-latitudes the 'classic' profile may be amended by a seasonal thermocline when surface waters are heated in spring and summer (Fig 3).

**13-27 Fig 3 Depth/Temperature Profiles for Different Latitudes**

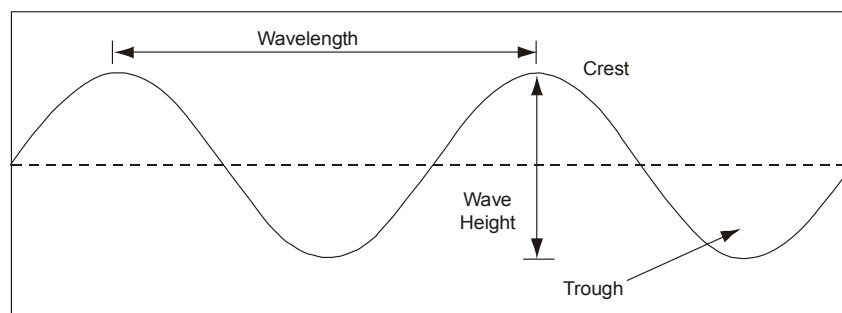


15. **Diurnal Effects.** Diurnal variations in temperature are insignificantly small, usually less than 0.3 °C in the open oceans although perhaps up to 3 °C in shallow coastal waters.

## Waves

16. **Formation.** Waves are another manifestation of the transfer of wind energy to the surface water. The precise mechanism of this transfer is complex and, as yet, poorly understood. In the area of formation, waves that are generated will consist of a superimposition of waveforms of varying frequency and amplitude depending upon the wind speed and the fetch (the length of sea area affected). Higher wind speeds lead to higher waves, but eventually a state of equilibrium will be reached with the excess energy being dissipated in, for example, white capping. The influence of waves is only felt close to the surface and waves generated by a severe storm would be virtually unnoticeable below about 100 m. The characteristics of waves are illustrated in Fig 4.

**13-27 Fig 4 Characteristics of an Ocean Wave**



Wave height - the linear distance between crest and trough

Wave length - the linear horizontal distance between corresponding points on consecutive waves

Wave period - the length of time for one complete cycle to pass a fixed point

17. **Swell.** Moving away from the area of generation the short period waves are dissipated, and only the long period waves remain. These are known as swell and can travel large distances from the generating area, typically thousands of kilometres.

18. **Sea Bed Effects.** When waves enter shallow water, they are affected by the seabed once the depth is less than  $\frac{1}{2}$  the wavelength. The front of the wave becomes steeper and eventually breaks.

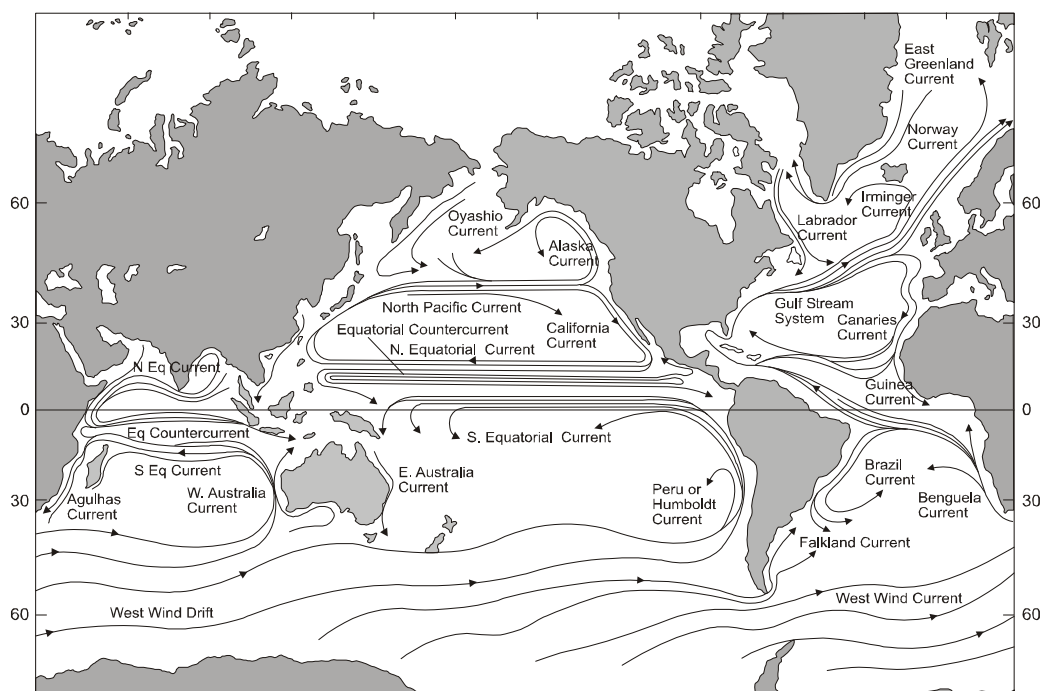
19. **Beaufort Scale.** The relationship between wind speed and sea state is expressed in the Beaufort Scale which may be used for estimating wind speed on land and at sea. It should be noted that, at sea, the scale is valid only for waves generated locally, and providing adequate time has elapsed and there has been an adequate fetch for a fully developed sea to become established. The Beaufort Scale is explained in Table 1 at the end of the Chapter.

## Currents

20. **Deep Currents.** As has been seen, sea water is not homogenous - its temperature may vary considerably, and there may be minor variations in salinity. In the Antarctic and around Greenland large masses of water are generated which are cold due to the interaction with the atmosphere, saline due to the freezing-out of a proportion of the fresh water, and therefore somewhat denser than average. This cold, dense water sinks and moves towards and across the Equator at depth. Clearly there will be boundaries between this water and the other water masses that it encounters with different temperature and salinity characteristics. Similar boundaries exist between other water masses with different properties. Examples are to be found where water from the Baltic enters the North Sea and where Mediterranean water enters the Atlantic Ocean. These boundaries are known as fronts and are analogous to meteorological fronts where different air masses meet.

21. **Surface Currents.** Sea currents are generated by winds when some of the energy of air movement is transferred by friction to the ocean surface. This leads to generalized zonal currents similar to the idealized coriolis-induced global patterns of surface winds illustrated in meteorological texts, and although this simple model is considerably modified by the presence of land masses, the influence of these winds is clearly discernable in Fig 5. As at depth there will be fronts between water masses of differing characteristics, reflecting their different origins and histories. Whereas some of these fronts will be more or less permanent, others will be of a temporary nature primarily due to seasonal effects.

13-27 Fig 5 Generalized Surface Currents



## The Biological Environment

22. **Life Forms.** Ninety-eight percent of marine species belong to the benthic, or seabed-living system. But, it is the remaining 2% of species, the pelagic or free-swimming system, which is of concern here. The life forms can be divided into two main groups: the plankton and the nekton (there is a third small group, the pleuston, which includes the jellyfish, which live at the sea surface, but they are of no relevance here).

23. **The Plankton.** The plankton are the group of organisms, both plant and animal, whose powers of locomotion are insufficient to prevent them being transported by currents. They range in size from less than 2  $\mu\text{m}$  up to just over 2 cms. As a large proportion of the plankton (the phytoplankton) is responsible for photosynthesis they must exist in that part of the ocean to which light can penetrate, typically the top 100 m or so. Virtually all plankton exhibit a diurnal variation of depth, rising near the surface at night and descending again just before sunrise.

24. **The Nekton.** The nekton are those creatures which can swim with sufficient power to be more or less independent of water movement. This group includes fish, cephalopods, and whales in the open ocean. Often they exhibit the same diurnal variation in depth as the plankton on which they feed.

25. **The Deep Scattering Layer.** The diurnal depth variation of most plankton and nekton has already been described and this event occurs within the top 100 m or so of the ocean. However, it is worth noting that there are some mainly carnivorous plankton and nekton which migrate between much greater depths. Typically, they will spend the day at between 400 and 1000 m and the night at about 200 to 300 m. The significance of this group of organisms is that they can affect echo sounders and SONAR equipment. It is thought that the reflection of sound waves is caused primarily by small fish possessing air bladders and by prawn-like creatures. This phenomenon is known as the deep scattering layer and can be observed in all oceans.



**Table 1 Sea State and Beaufort Wind Scale**

Sea State	Beaufort No	Name	Wind Speed		Characteristics	Wave Height (m)	Wind Gusts
			knots	ms <sup>-1</sup>			
0	0	Calm	1	0.0-0.2	Sea like mirror. Complete calm. Smoke rises vertically.	0	-
	1	Light Air	1-3	0.3-1.5	Ripples with appearance of scales, no foam crests. Wind direction shown by smoke drifts but not by wind vanes.	0.1-0.2	-
1	2	Light Breeze	4-6	1.6-3.3	Small wavelets, crests have glassy appearance but do not break. Wind felt in face, leaves rustle, ordinary vanes moved by wind.	0.3-0.5	-
2	3	Gentle Breeze	7-10	3.4-5.4	Large wavelets, crests begin to form, scattered white horses. Leaves and small twigs in constant motion, wind extends light flag.	0.6-1.0	-
3	4	Moderate Breeze	11-16	5.5-7.9	Small waves becoming longer, fairly frequent white horses. Wind raises dust and loose paper, small branches are moved.	1.5	-
4	5	Fresh Breeze	17-21	8.0-10.7	At sea, moderate waves assume longer form, many white horses and chance of spray, crested wavelets on inland lakes. Small trees in leaf begin to sway.	2.0	-
5	6	Strong Breeze	22-27	10.8-13.8	Large waves, extensive white foam crests, spray probable. Large branches in motion, umbrellas used with difficulty, whistling in telegraph wires.	3.5	-
6	7	Near Gale	28-33	13.9-17.1	Sea heaps up, white foam begins to be blown in streaks. All trees in motion, inconvenience felt when walking against wind.	5.0	-
7	8	Gale	34-40	17.2-20.7	Moderately high waves of greater length, crests break into spindrift, foam well blown in streaks. Twigs break off, wind impedes progress.	7.5	43-51
8	9	Strong or Severe Gale	41-47	20.8-24.4	High waves, dense streaks of foam, spray may affect visibility, sea begins to roll. Slight structural damage occurs (chimney pots and slates removed).	9.5	52-60
9	10	Storm	48-55	24.5-28.4	Very high waves, overhanging crests, sea surface white, heavy rolling and visibility reduced. Trees uprooted, considerable structural damage occurs.	12.0	61-68

	11	Violent Storm	56-64	28.5-32.7	Exceptionally high waves, small and medium sized ships lost to view in troughs. Widespread wind damage on land.	15.0	69-77
	12	Hurricane	64+	32.7+	Air filled with foam and spray, sea completely white with driving spray. Visibility greatly reduced. Structural damage to disaster level.	15+	78+

## CHAPTER 28 - SOUND IN THE SEA

Contents	Page
Introduction .....	2
The Velocity of Sound in the Sea .....	2
Refraction .....	3
Shadow Zones.....	5
Sound Transmission Modes (Paths) .....	5
Leakage Paths.....	10
Coherency Effect (Lloyd's Mirror).....	11
Diurnal (Afternoon) Effect.....	11
Noise.....	11
Sound Transmission Losses .....	12

### Table of Figures

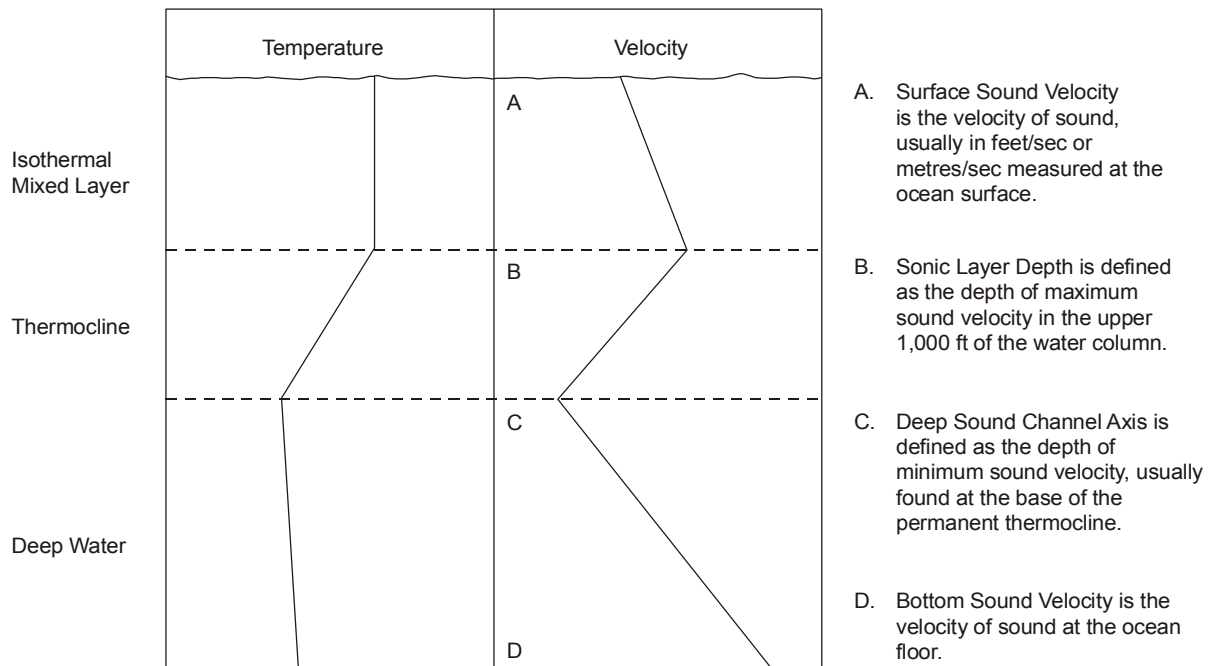
13-28 Fig 1 Sound Velocity Profile for a Typical 'Three-layer Ocean'.....	3
13-28 Fig 2 Refraction in a Layered Medium .....	4
13-28 Fig 3 Velocity Gradients .....	6
13-28 Fig 4 Surface Duct .....	7
13-28 Fig 5 Sound Channel .....	8
13-28 Fig 6 Convergence Zone Propagation .....	9
13-28 Fig 7 Convergence Zone Velocity Profile.....	9
13-28 Fig 8 Leakage Effects .....	10
13-28 Fig 9 Lloyd's Mirror Effect .....	11
13-28 Fig 10 Spreading Loss Graph .....	13

## Introduction

1. Due to the rapid attenuation of other energy forms in the sea, e.g. light, the use of sound is one of the few methods available for underwater detection. A study of the behaviour of sound in the sea is therefore important for operators of maritime patrol aircraft.
2. Unlike light, which is a form of electromagnetic energy, sound involves the vibration of the medium through which it is transmitted. So whereas light can pass through a vacuum, sound cannot, generally travelling best through solids and liquids and somewhat less well through gases.
3. The wavelengths of sound which are relevant in the ocean range from about 50 metres to 1 millimetre. Assuming a velocity of sound in sea water of  $1,500 \text{ ms}^{-1}$ , this corresponds to frequencies between 30 Hz and 1.5 MHz. Variations in the velocity of sound in sea water and their effect on propagation are in fact very important and will be examined in some detail.

## The Velocity of Sound in the Sea

4. In Volume 13, Chapter 25 a simple equation relating the velocity of sound in air to specific heat, pressure and density was given. Unfortunately, this simple relationship does not apply in liquids and although a mathematical treatment can be derived, it is rather complex. However, it is possible to deal empirically with the variation of sound velocity with salinity, pressure and temperature. An increase in any one factor will increase the sound velocity if all other factors remain constant. In line with current practice, changes are shown in units of feet, pounds per square inch (psi) and °F rather than in SI units. The velocity of sound in sea water varies between  $4,700 \text{ fts}^{-1}$  and  $5,100 \text{ fts}^{-1}$ .
5. **Salinity.** A 1 o/oo increase in salinity increases the sound velocity by  $4.27 \text{ fts}^{-1}$ . Unfortunately there is no method of readily determining salinity from a maritime patrol aircraft but assuming a constant value of 35 o/oo is normally sufficient. However, caution is needed in those areas where salinity values are liable to be markedly different from the average, e.g. in coastal and ice melt areas.
6. **Pressure.** An increase in pressure of 44 psi, corresponding to a depth change of about 100 ft, will increase the sound velocity by about  $1.7$  to  $1.8 \text{ fts}^{-1}$ . Fortunately pressure increases with depth in a linear fashion and so velocity changes are easily predicted with fair accuracy.
7. **Temperature.** Of the 3 factors being considered, changes in temperature have the greatest effect on sound velocity. A  $0.5 \text{ }^{\circ}\text{C}$  ( $\equiv 1 \text{ }^{\circ}\text{F}$ ) increase in water temperature will cause an increase in sound velocity of between 4 and  $8 \text{ fts}^{-1}$ .
8. **Sound Velocity Profile.** Fig 1 shows how sound velocity varies with depth in a typical 'Three-layer Ocean'. This variation with depth is known as a sound velocity profile (SVP). The following points can be deduced:
  - a. There is an increasing velocity down through the surface mixed layer due to the effect of increasing pressure (Temperature constant).
  - b. There is decreasing velocity with depth in the thermocline due to decreasing temperature having a greater effect than increasing pressure.
  - c. There is increasing velocity down through the deep water due to increasing pressure (temperature constant).

**13-28 Fig 1 Sound Velocity Profile for a Typical 'Three-layer Ocean'**

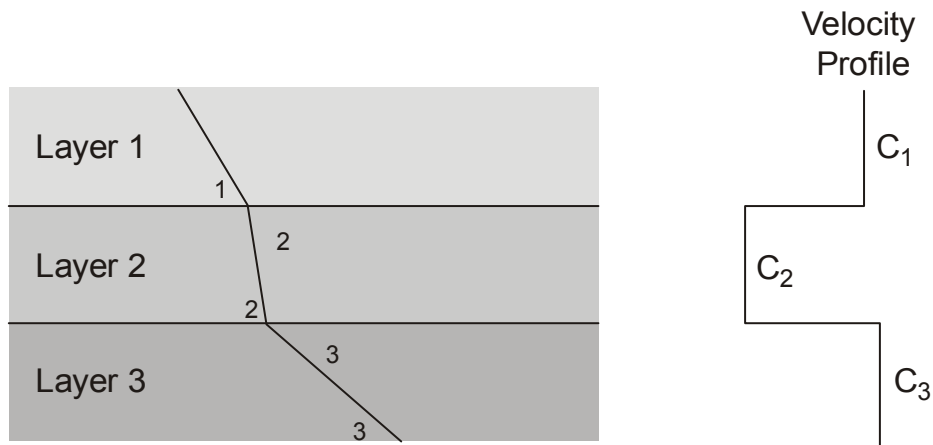
9. **Temperature Gradient and Sound Velocity.** Temperature/depth profiles are readily obtained by MPA crews and the following relationship between temperature profile and sound velocity can be expected:

- a. If temperature decreases with depth at a rate of 0.1 to 0.2 °C/100 ft (0.2 to 0.4 °F/100 ft) its effect will be offset almost exactly by the effect of the increasing pressure with depth. This will result in an ISOVELOCITY condition (i.e. no change in velocity with depth).
- b. If temperature decreases with depth at a rate of less than 0.1 to 0.2 °C/100 ft (0.2 to 0.4 °F/100 ft) then sound velocity will increase with depth and a POSITIVE velocity gradient will exist.
- c. If temperature decreases with depth at a rate greater than 0.1 to 0.2 °C/100 ft (0.2 to 0.4 °F/100 ft) then sound velocity will decrease with depth and a NEGATIVE velocity gradient will exist.
- d. If temperature remains constant with depth, this is termed ISOTHERMAL and there will be velocity gradient of +2 (i.e. an increase in velocity of 2  $\text{fts}^{-1}$  per 100 ft increase in depth).

## Refraction

10. When sound leaves a source it can be considered to move along paths known as rays - analogous to light rays. As with light (see Volume 13, Chapter 21, Para 21), if the medium of transmission is homogenous, then these rays will be straight lines but if the ray moves into an area where the velocity of sound is different then the ray will be bent, a phenomenon known as refraction.

11. The effect of refraction on a ray passing through layers of different velocity is illustrated in Fig 2 where the total ray path will be seen to be made up of a series of straight-line segments.

**13-28 Fig 2 Refraction in a Layered Medium**

12. A useful mnemonic is the 'HALT' rule which states that rays will be refracted away from depths of relatively high velocity and towards depths of relatively low velocity:

**High Away, Low Towards**

Thus in Fig 2 the ray moving from layer 1 to layer 2 (relatively low velocity) will be refracted towards layer 2 i.e.  $\theta_2 > \theta_1$ . On reaching layer 3 (relatively high velocity) the ray will be refracted away from layer 3 i.e.  $\theta_3 < \theta_2$ .

13. Applying this rule to the typical SVP in Fig 1, it will be seen that sound rays in the mixed surface layer will be refracted upwards away from the Sonic Layer Depth and sound rays in the thermocline will be refracted downwards until they pass the depth of the Deep Sound Channel Axis when they will be refracted upwards again. The direction of propagation of a sound ray will therefore be determined by the ambient SVP.

14. **Snell's Law.** The degree to which a ray is bent when moving from an area of one velocity to another is determined by Snell's Law which states that:

In a medium comprising discrete layers each of different sound velocity, the angles  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  etc of a ray incident on and leaving a boundary between layers are related to the sound velocities,  $C_1$ ,  $C_2$ ,  $C_3$  etc of these respective layers such that:

$$\frac{\cos \theta_1}{C_1} = \frac{\cos \theta_2}{C_2} = \frac{\cos \theta_3}{C_3} = \text{a constant, for any given ray.}$$

15. **Limiting Ray.** When sound rays are refracted within a sea layer, one ray will eventually just graze or be tangential to the boundary with an adjacent layer, or to the sea surface or ocean bottom. This ray, which will continue to be refracted within the layer, is known as the limiting ray. Rays which approach the layer boundary at a more perpendicular angle, will pass through it (or be reflected or absorbed).

## Shadow Zones

16. The HALT rule predicts that rays will be refracted away from depths of sound velocity maxima. As a result there often exists at these depths a region into which very little acoustic energy can penetrate; such regions are termed 'Shadow Zones'. The existence of such a zone, and its relative position, depends upon the SVP. Minor amounts of energy will enter the shadow zone due to diffractive and scattering effects. The nature of shadow zones in a variety of SVPs is shown in Fig 3 a-d.

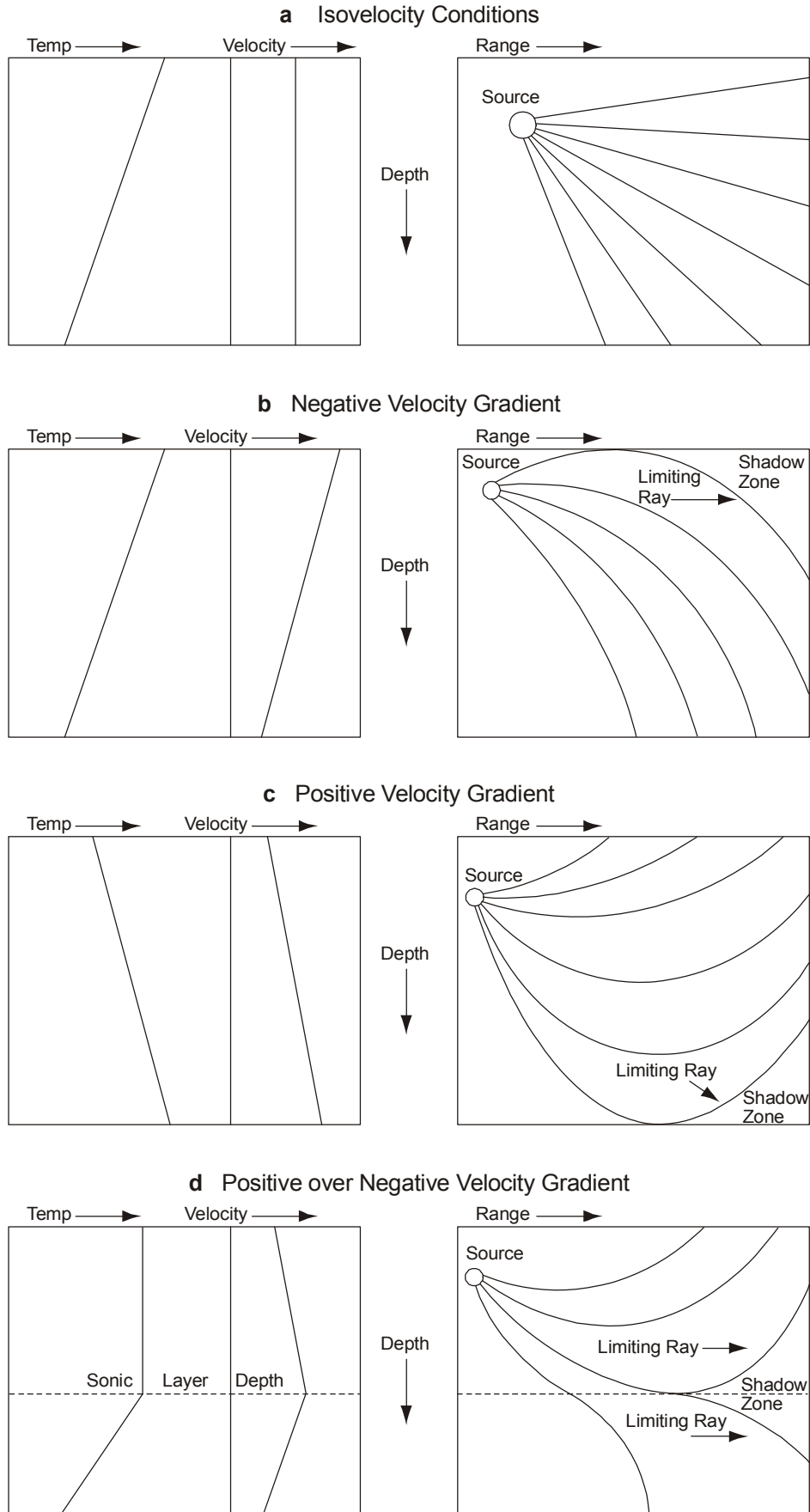
- a. **Isovelocity Conditions (Fig 3a).** In isovelocity conditions, the increase in sound velocity due to increasing pressure with depth is offset by the effect of decreasing temperature with depth. Since there is no variation in velocity there is no refraction and sound rays travel in straight lines from the source.
- b. **Negative Velocity Gradient (Fig 3b).** In this situation, the temperature decreases at a rate sufficient to overcome the pressure effect and so velocity decreases with depth. Sound rays are refracted downwards away from the higher velocity area. A shadow zone exists above and beyond the limiting ray.
- c. **Positive Velocity Gradient (Fig 3c).** In the situation where temperature is constant or increasing with depth, or at least does not fall at a rate to override the pressure effect, then a positive velocity gradient will be established. Sound rays will be refracted upwards away from the higher velocity area and a shadow zone exists below and beyond the limiting ray.
- d. **Positive over Negative Velocity Gradient (Fig 3d).** The combination of a positive velocity gradient above a negative velocity gradient produces a split beam effect at the depth of the velocity gradient change. The limiting ray and all rays above it are refracted upwards away from the higher velocity area. All rays below the limiting ray are similarly refracted downwards. A shadow zone will exist beyond the limiting ray(s).

## Sound Transmission Modes (Paths)

17. Clearly sound energy can travel directly underwater from the source to the detector, affected only by 'bending' due to refraction. In general, however, direct rays achieve only relatively short ranges. The range is normally determined by the limiting ray and will depend primarily on the Sonic Layer Depth, the SVP below the surface mixed layer, and the source/receiver depth combination:

- a. **Effect of Sonic Layer Depth.** The deeper the Sonic Layer Depth, the greater will be the Direct Path range. This is because the limiting ray has a greater vertical distance to travel before being refracted upwards and those rays at a greater angle than the critical angle ray will also have a greater vertical distance to travel before passing through the layer.
- b. **Effect of SVP Below the Surface Mixed Layer.** The greater the negative velocity gradient below the surface mixed layer, the more pronounced will be the refraction and so the more reduced will be the direct path range.
- c. **Effect of Receiver/Source Depth Combination.** If the receiver and source are in different layers then this will curtail the direct path range to an extent, particularly if the receiver is shallower than the source. The greatest direct path range is normally achieved when both source and receiver are in an isothermal layer. In general though, the direct path range depends upon the velocity gradient.

**13-28 Fig 3 Velocity Gradients**



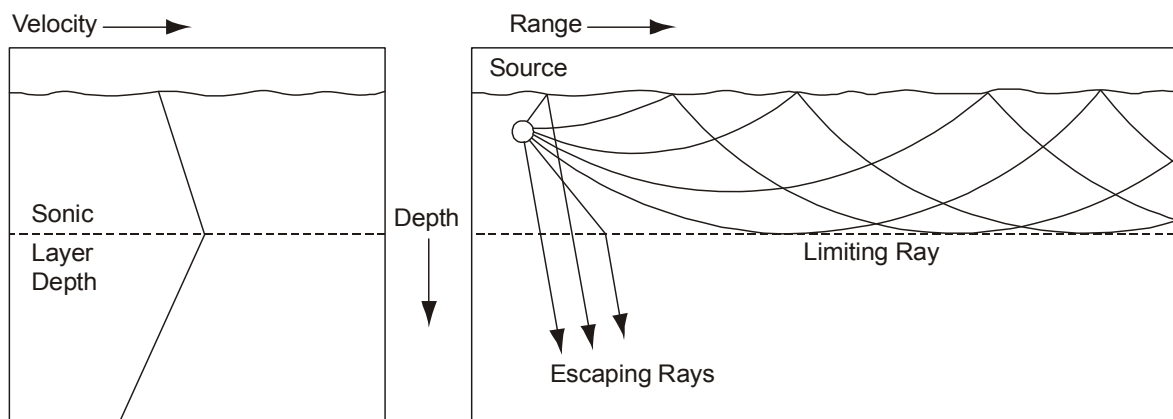


18. **Phenomena.** The variations in sound velocity with depth in the ocean lead to a number of effects which allow much greater sound detection ranges to be achieved than would be possible if the direct path only was available. The infinite variety of sound velocity profiles implies an infinite variety of transmission modes. However it is possible to group them all into four basic phenomena; other modes tend to be just variations on these four. It should be remembered though, that unless transmitter and receiver are very close together, then the received sound has very probably been transmitted by a combination of more than one mode.

19. **Surface Duct (Surface Sound Channel).** This transmission mode, sometimes known as 'The Mixed Layer Sound Channel', concerns only acoustic energy contained within the surface layer. The prime requisite for the existence of a surface duct is a layer with a positive velocity gradient. Sound rays in the layer are refracted up towards the surface (away from the higher velocity), reflected off the sea surface downwards and are then refracted upwards again and so on. This trapping of the sound energy within the duct can result in extended ranges. The quality of sound transmission via this mode is dependent upon the Sonic Layer depth, the sea state, and the source depth. Whether this mode can be used depends primarily upon whether the sensor can be placed in the layer and also upon the signal frequency. Surface Duct propagation is illustrated in Fig 4, and its usefulness may be affected by the following factors:

- a. **Sonic Layer Depth.** An effective Surface Duct cannot exist unless the Sonic Layer depth is greater than 50 feet. The deeper the Sonic Layer depth the greater the range in the duct as the sound rays have further to travel before being reflected off the surface. Each time a sound ray is reflected off the surface some energy is lost by scattering so for any given range the reflection losses are less for a deep layer.
- b. **Sea State.** Wind, waves and swell all increase absorption, reflection and scattering losses as well as increasing the ambient noise by inducing air bubbles into the water. These factors will all reduce the range achieved and the effectiveness of the duct.
- c. **Velocity Gradient in the Surface Layer.** In order to have an effective surface duct the sound velocity gradient within the surface layer must be positive. The more positive the gradient the less sound energy that can escape and the better the detection ranges. Isovelocity conditions represent the limit of a positive gradient and result in straight-line propagation, the limiting type of Surface Duct.

13-28 Fig 4 Surface Duct

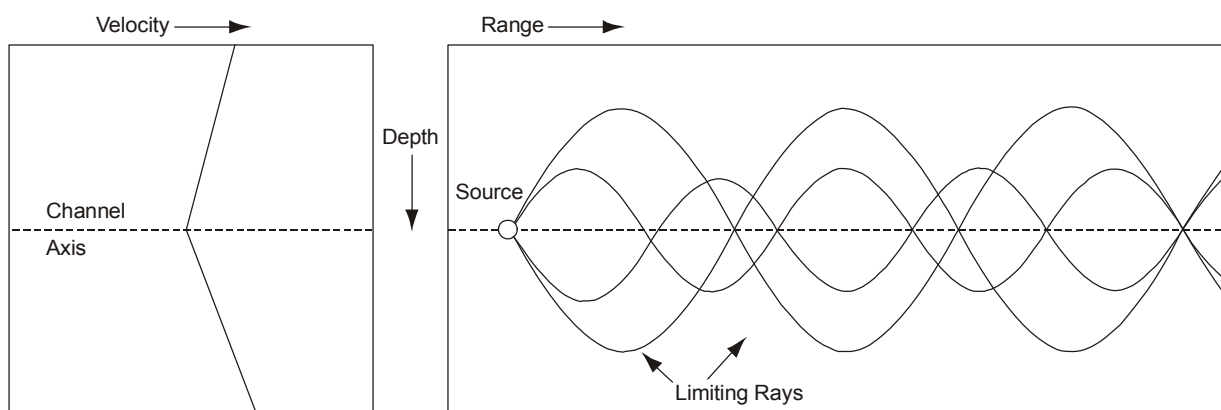


20. **Sound Channels.** The pre-requisite for a Sound Channel is a region of decreasing sound velocity overlaying a region of increasing sound velocity. Applying the HALT rule will show that sound rays will tend to be refracted towards the depth of minimum sound velocity. This depth is known as the Sound Channel Axis. A Sound Channel is illustrated in Fig 5. The following two types of sound channel are recognized:

- a. **Deep Sound Channel.** This is also known as the SOFAR Channel (Sound Fixing And Ranging), and is the most common type of Sound Channel. It results from the negative velocity gradient of the thermocline overlaying the positive velocity gradient established by the pressure effect in deep water. The axis of the Deep Sound Channel is found at the depth of minimum sound velocity, usually at the base of the permanent thermocline. The SOFAR axis depth varies geographically. In the Atlantic Ocean, it varies between 1,300 ft in high latitudes (above 60° N) and 6,600 ft in low latitudes (about 35° N). The depth also decreases uniformly with longitude away from the Greenwich Meridian in the North Atlantic.
- b. **Shallow Sound Channel.** This is sometimes known as a Depressed Sound Channel or Sub-Surface Duct. This type of sound channel occurs in the upper layers of water above the permanent thermocline but is somewhat transient in nature. In order to be effective the axis depth of a Shallow Sound Channel must be between 150 and 600 feet below the surface. The channel must be at least 50 feet thick and the limiting rays must be at an angle of at least 2° above and below the channel axis. Shallow Sound Channels exist in the Mediterranean and the North-East Atlantic at certain times of the year and are present in the lower latitudes of the Atlantic all year. In the North Atlantic, they are only established during the summer months, with an axis depth of about 450 feet.

The usability of both types of sound channel is determined by the proximity of the receiver to the axis depth and by the signal frequency.

13-28 Fig 5 Sound Channel



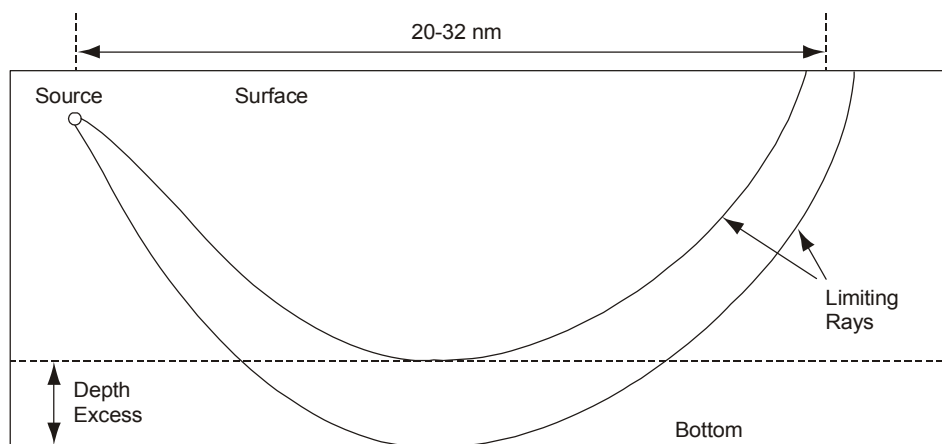
21. **Convergence Zone.** Sound rays leaving a source near the surface which penetrate below the surface layer, will be refracted downwards until they pass the depth of the deep sound channel axis whence they begin to be refracted upwards towards the surface. Provided that the water depth is great enough a sufficient number of rays will eventually reach the near surface depths as a ring of focused energy around the source known as the annulus. This is illustrated in Fig 6. The likelihood of

Convergence Zone propagation can be determined from a sound velocity profile using the following parameters which are shown graphically in Fig 7:

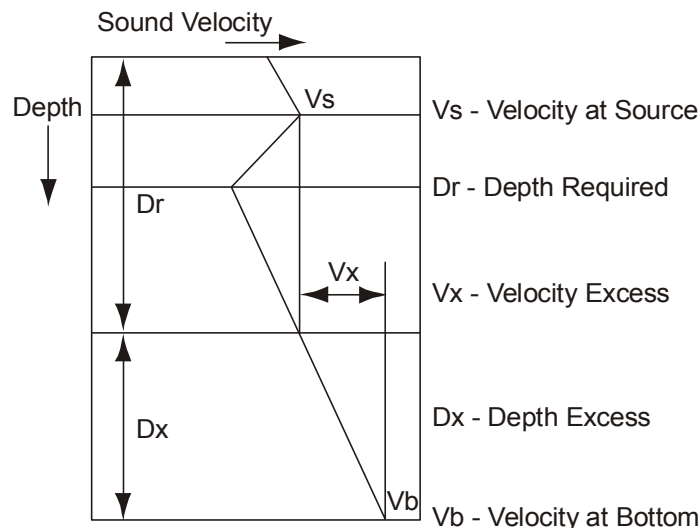
- Velocity at Source ( $V_s$ )**. This is the sound velocity at the source depth.
- Velocity at Bottom ( $V_b$ )**. This is the sound velocity at the ocean bottom depth and is used to determine whether or not sound rays will be refracted back upwards towards the surface. Before any rays will be refracted upwards, the velocity at the bottom ( $V_b$ ) must exceed the sound velocity at the source depth ( $V_s$ ).
- Velocity Excess ( $V_x$ )**. If  $V_b$  is greater than  $V_s$ , then the difference is known as the 'Velocity Excess'. For a reliable Convergence Zone, this velocity excess must be approximately  $33 \text{ fts}^{-1}$ .

Convergence Zone range in the UK MPA's normal operating areas is between 20 and 32 nautical miles. It is unusual for Convergence Zone propagation to take place in warm or moderately warm water in depths of less than 1,200 fathoms. However, Convergence Zone can exist in water depths of less than 300 fathoms in certain conditions. Convergence zone width is directly proportional to the depth excess and can be roughly estimated at 10% of the range interval. It is possible to have a number of Convergence Zones around a single source. The first zone will be the strongest in terms of intensity. The second and third zones occur, respectively, at twice and three times the range of the first, providing that the source is strong enough.

**13-28 Fig 6 Convergence Zone Propagation**



**13-28 Fig 7 Convergence Zone Velocity Profile**



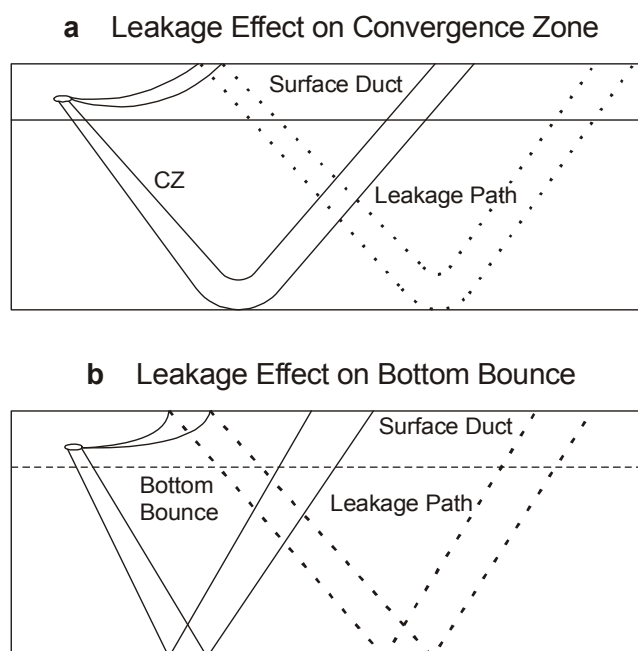
22. **Bottom Bounce Mode.** Sound rays which leave a source at angles greater than that of the Limiting Ray will eventually strike either the ocean bottom or the sea surface. The 'Bottom Bounce Mode' refers to those rays which are reflected back and forth between these two boundaries. Whereas some energy will be reflected off the boundary surface (either ocean bottom or sea surface), a proportion will be lost to sensors due to scattering and absorption. The quality of this transmission mode is primarily dependent on factors such as boundary surface roughness, bottom type, water depth, and signal frequency. Bottom Bounce Mode is particularly important in shallow water or where the Sonic Layer depth is less than 50 feet. In fact, sound propagation for frequencies between 50 Hz and 1,500 Hz in shallow water is dominated by the 'Bottom Bounce' transmission mode.

### Leakage Paths

23. In general sound channels or ducts are low loss propagation paths. At low frequencies, however, it is found that losses increase suddenly and the duct 'breaks down'. The frequency at which this effect occurs is known as the cut-off frequency. In simple terms the wavelength becomes too big to fit into the duct. The leakage path is in the direction of propagation and the main cause of the leakage is diffraction together with some scattering losses. In Volume 13, Chapter 25, when diffraction was being considered, it was shown that according to Huygens' principle secondary wavelets are produced from each point on a wave-front. Although the vast majority of the sound energy is propagated in the forward direction, these wavelets have a component of their energy to the side and this energy can 'leak' into shadow zones or out of a sound channel. Leakage paths tend to complicate the relatively simple propagation paths that have been outlined. Two examples are:

- a. **Leakage Effect on Convergence Zone.** Fig 8a illustrates the effect of leakage on convergence zone propagation. Whereas some energy will enter the convergence zone path immediately, some may remain trapped in the surface duct for some range before leaking into a different convergence zone path. This anomalous propagation can lead to confusion in recognizing the convergence zone and the width of the annulus.
- b. **Leakage Effect on Bottom Bounce.** Fig 8b illustrates the effect of leakage on bottom bounce propagation. Again some energy is transmitted directly to the bottom while some remains trapped in the surface duct for a while before entering a bottom bounce path.

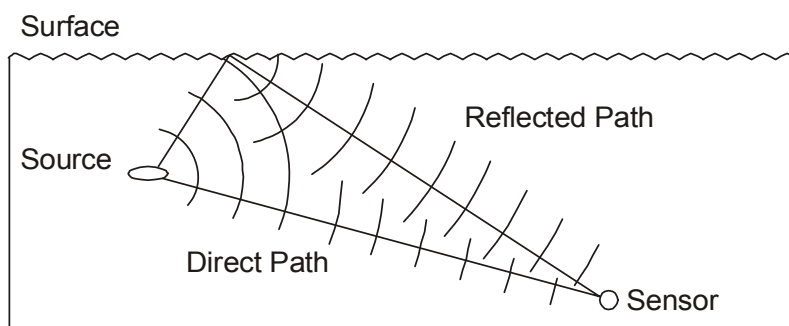
13-28 Fig 8 Leakage Effects



### Coherency Effect (Lloyd's Mirror)

24. Clearly, it is possible for sound from one source to arrive at a receiver having followed a variety of paths. In general, any two sound paths will be of different lengths and so it is highly likely that the phase of the sound arriving by one path will be different to that of the sound arriving by another. If the received sound waves are in phase then there will be an enhancement of the signal. If, however, the phases are different then there will be a reduction in the received signal with complete cancellation if the waves are  $180^\circ$  out of phase. Commonly this effect is a result of interference between a direct wave and a surface reflected wave or, in shallow water, between a direct wave and a bottom-reflected wave. A similar effect can occur if some sound is reflected by an object in the water and therefore arrives out of phase with the direct wave (secondary source polarisation). Lloyd's mirror effect is illustrated in Fig 9.

13-28 Fig 9 Lloyd's Mirror Effect



### Diurnal (Afternoon) Effect

25. Solar heating near the surface water can cause short-term variations in the vertical temperature structure which can, in turn, bring about changes in sound propagation paths. These diurnal variations can lead to shortening of detection ranges. As the surface water is heated during the day, a negative temperature profile builds up causing refraction of the sound energy and the establishment of new shadow zones. As the surface water cools again, the temperature gradient and its associated effects disappear.

### Noise

26. **Ambient Noise.** Ambient noise is that unwanted sound which is inherent in the sea independent of either the target or the aircraft. It can, of course, mask the noise generated by a target which the sensors are attempting to detect. Ambient noise can be predicted by graphical or computer methods (both of which depend upon a mix of theoretical and empirical data), or it can be measured using an ambient noise meter. There are three main sources of ambient noise in the areas where the UK MPA normally operate:

- a. **Shipping Traffic.** Shipping tends to generate the main component of ambient noise in deep-water areas, especially at lower frequencies (below 500 Hz). It is most intense in the vicinity of major shipping lanes and results from the sound of machinery, electrical systems, hydraulics, hydrodynamic flow, and propeller cavitation.
- b. **Sea State.** This noise source is the result of sea surface agitation causing the entrapment and subsequent escape of air bubbles. This noise is dominant at higher frequencies (above 500 Hz). Sea state noise is increased by strong surface winds especially in shallow water.
- c. **Biological.** Some crustaceans, fish and marine mammals contribute to ambient noise by producing sounds associated with their normal life activities. Biological sounds are present over

a large frequency spectrum and are extremely difficult to predict with any accuracy due to such factors as changing biological cycles and movement of the source organisms; there are often seasonal and diurnal cycles. In general, biological noise is greatest in shallow water, particularly in tropical and sub-tropical seas.

27. **Self-noise.** Self-noise is a noise generated by a maritime patrol aircraft's own mechanical, avionic, and electrical systems.

### Sound Transmission Losses

28. Not all of the sound generated by a source reaches a sensor. There are two general ways in which acoustic energy may be dissipated during propagation: spreading and attenuation.

29. **Spreading Losses.** Spreading is a geometrical phenomenon whereby a fixed amount of energy is distributed over an ever-increasing area as it moves away from the source. There are two types of spreading loss which may be encountered.

a. **Spherical Spreading.** If the sound velocity in the ocean is constant in both the horizontal and vertical directions, then the sound energy will radiate equally in all directions from the source. At any given distance,  $r$ , from the source, the power of the source,  $W$ , will be spread over the surface of a sphere of radius  $r$ . This surface area is  $4\pi r^2$ . The sound intensity,  $I$ , which is the sound power per unit area, will be given by:

$$I = \frac{W}{4\pi r^2} \text{ Wm}^{-2}$$

This is an inverse square law, i.e. intensity is inversely proportional to the square of the range. Notice that it is frequency independent. Working in Sound Pressure Levels (SPLs) in dBs, then the spreading loss (spherical) =  $20 \log r$ . If the range is doubled from  $r$  to  $2r$  then:

$$\begin{aligned} \text{Change in SPL} &= 20 \log r - 20 \log 2r \\ &= 20 \log [r \div 2r] \\ &= 20 \log 0.5 = 20 \times -0.3 \\ &= -6 \text{ dB} \end{aligned}$$

i.e. a doubling of distance results in a 6dB loss.

b. **Cylindrical Spreading.** Cylindrical spreading occurs when the acoustic energy is trapped between two parallel boundaries such as the upper and lower bounds of a sound channel, surface duct or a convergence zone. The energy is constrained to spread in only two dimensions in a cylindrical manner. In this case, at any given distance from the source,  $r$ , the power of the source,  $W$ , will be spread over the surface of a cylinder of radius  $r$ . The area in this case will be  $2\pi r$  and the variation of sound intensity with range is given by:

$$I = \frac{W}{2\pi r} \text{ Wm}^{-2}$$

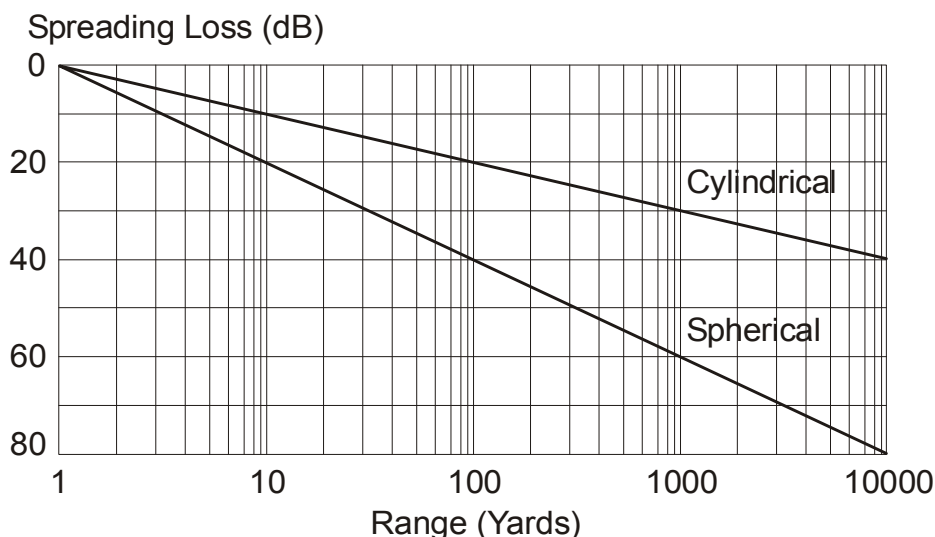
The spreading loss (cylindrical) in dBs =  $10 \log r$ , so looking again at the loss involved in a doubling of range:

$$\begin{aligned} \text{Change in SPL} &= 10 \log r - 10 \log 2r \\ &= 10 \log [r \div 2r] \\ &= 10 \log 0.5 = 10 \times -0.3 \\ &= -3 \text{ dB} \end{aligned}$$

i.e. doubling of distance results in a 3dB loss.

30. **Spreading Loss Comparisons.** Fig 10 shows a comparison between the different types of spreading loss. Since cylindrical loss is less severe than spherical loss, ducted modes of sound propagation often yield the highest probabilities of signal detection. However, most of the time the received signal has suffered some combination of the spreading losses. For example, since direct path makes up the initial portion of ALL transmission modes, they all exhibit some degree of spherical spreading.

13-28 Fig 10 Spreading Loss Graph



31. **Attenuation Losses.** Attenuation is the term applied to the linear decrease in acoustic energy per unit area of a wave-front as the distance from the source increases. Attenuation loss includes the effects of absorption, scattering and diffractive leakage.

a. **Absorption Loss.** Absorption loss involves the conversion of acoustic energy into heat by molecular action. Absorption loss is directly proportional to the range from the source. Increases in salinity and decreases in temperature increase absorption losses. Unlike spreading losses, absorption depends upon the frequency, varying approximately as the square of the frequency.

b. **Scattering Loss.** Scattering is the random reflection of acoustic energy from the ocean surface, the ocean bottom, or from suspended particles (volume scattering). Factors influencing the degree of each type of scattering are:

(1) **Surface Scattering.** The severity of surface scattering loss is dependent upon wave height, signal frequency, and angle of incidence of the sound energy.

(2) **Bottom Scattering.** The severity of bottom scattering is dependent upon the bottom roughness, sediment particle size, signal frequency, and angle of incidence of the sound energy.

(3) **Volume Scattering.** The severity of the scattering loss due to volume scattering is the most difficult to predict. It is dependent upon the ratio of the particle size responsible for the reflection to the signal wavelength. It is also dependent on the type of particle (i.e. solid or fluid). The most important contributor to volume scattering is biological in nature. Apart from the effect of the Deep Scattering Layer, volume scattering strength tends to decrease with depth.

The prediction of actual losses due to scattering is problematical, but there are some general observations that can be made: higher scattering losses are associated with higher frequencies, rougher scattering surfaces and larger scattering particles. Fluids such as bubbles are generally more effective volume scatterers than solid particles.

## CHAPTER 29 - STATICS

Contents	Page
Introduction .....	1
Forces, Moments and Couples .....	1
Centre of Gravity .....	2
Equilibrium .....	3
Stability .....	4
Friction .....	4
Machines .....	5

### Table of Figures

13-29 Fig 1 Determining the Centre of Gravity of a Flat Plate .....	2
13-29 Fig 2 Determining the Centre of Gravity of a Symmetrical Body .....	2
13-29 Fig 3 Translational Equilibrium .....	3
13-29 Fig 4 Rotational Equilibrium .....	4
13-29 Fig 5 Lever in Rotational Equilibrium .....	5
13-29 Fig 6 Lever - Smaller Force over Greater Distance .....	6
13-29 Fig 7 A Two-pulley System .....	6

### Introduction

1. Statics is the study of forces in equilibrium. A particle is in equilibrium when all the forces acting upon it are balanced; it may be stationary or it may be in a state of unchanging motion. The term force and some associated terms are discussed below.

### Forces, Moments and Couples

2. **Force.** A force is that quantity which when acting on a body which is free to move, produces an acceleration in the motion of that body. For example, if a stationary football is kicked, it moves, ie it experiences an acceleration; the kick is the applied force. If a moving football is kicked, it changes speed, or direction, or both, so again it experiences an acceleration. However, just because a body is not accelerating does not mean that there are no forces acting upon it. Although that may be the case, it may equally be the situation that there are several forces acting such that their resultant is zero. For example, an aircraft which is flying at constant altitude, speed, and direction has many forces acting upon it, principally lift, thrust, weight, and drag.

3. **Weight.** Any body on or near the Earth is subject to the attraction of Earth's gravity, and if a body is released in this situation it will accelerate towards the centre of the Earth. The value of the acceleration, which is usually given the symbol  $g$ , is approximately  $9.8 \text{ ms}^{-2}$  but varies slightly both geographically and with altitude. Since gravity imposes an acceleration on a body, it must be a force, and it is this force which is known as weight.

4. **Vector Representation of a Force.** Force is a vector quantity, ie it has both magnitude and direction. A vector may be represented by a straight line drawn to scale and marked with an arrow. The direction of the line represents the direction of the vector and its length represents the magnitude. Vectors may be added together by means of the triangle, parallelogram, and polygon laws to give a single resultant and a single vector may be resolved into two or more components. A more complete treatment of vectors is to be found in Volume 13, Chapter 4.



5. **Moment or Torque.** The moment of a force about a point, or torque, is the tendency of the force to turn the body to which it is applied about that point. The magnitude of the moment or torque is the product of the force and the perpendicular distance from the point to the line of the force. The moment of a force about an axis is the product of the force and the length of the perpendicular common to the line of the force and the axis.

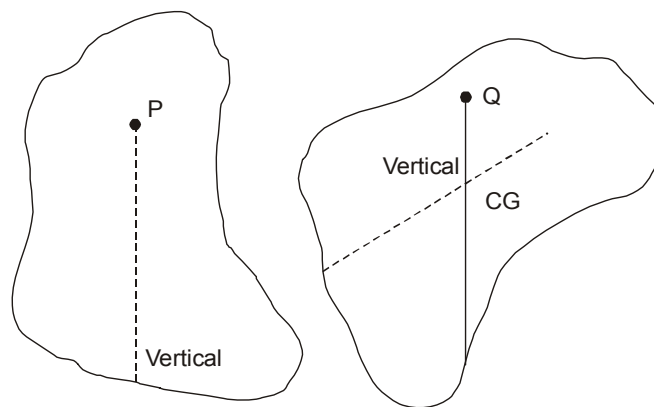
6. **Couple.** A system of two equal and parallel forces, acting in opposite directions but not in the same line, is known as a couple. The moment of a couple about any point in the plane of the forces is constant and equal to the product of one of the forces and the perpendicular distance between their lines of action.

### Centre of Gravity

7. In any rigid extended body there is a unique point at which the total gravitational force, the weight, appears to act. This point is known as the centre of gravity.

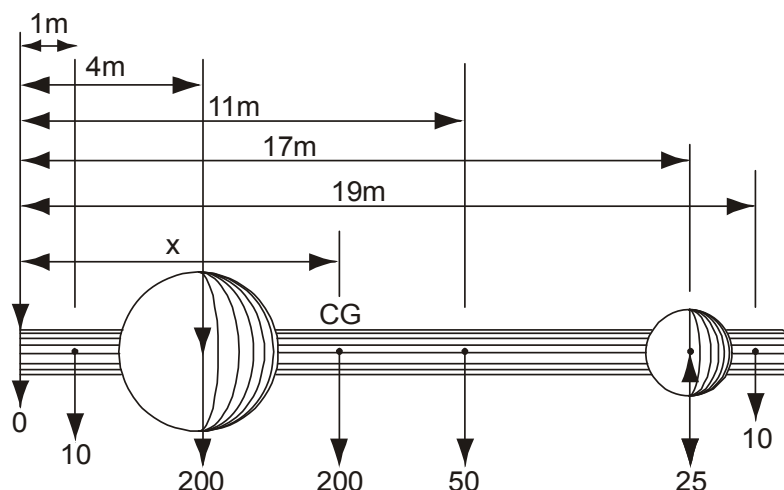
8. The position of the centre of gravity of a flat body can be determined by suspending it at any point, P and marking the vertical, then suspending it at a second point, Q, and again marking the vertical. The centre of gravity is at the intersection of the two lines (Fig 1).

**13-29 Fig 1 Determining the Centre of Gravity of a Flat Plate**



9. The position of the centre of gravity of a composite body, which can be treated as a number of symmetrical parts, can be found as shown in the following example. A body consists of two uniform spheres mounted on a uniform cylindrical bar as shown in Fig 2.

**13-29 Fig 2 Determining the Centre of Gravity of a Symmetrical Body**



Taking 0 as the reference point, the parts can be treated as three cylinders with centres of gravity distances 1, 11, and 19, metres respectively from the datum, and weights 10, 50, and 10 units; and two spheres of weights 200 and 25 units, with centres of gravity 4 and 17 metres respectively from the datum. Let the centre of gravity, through which the total weight (295 units) of the body acts, be  $x$  metres from 0. Taking moments about 0:

$$(10 \times 1) + (200 \times 4) + (50 \times 11) + (25 \times 17) + (10 \times 19) = 295x$$

$$\therefore 295x = 10 + 800 + 550 + 425 + 190 = 1975$$

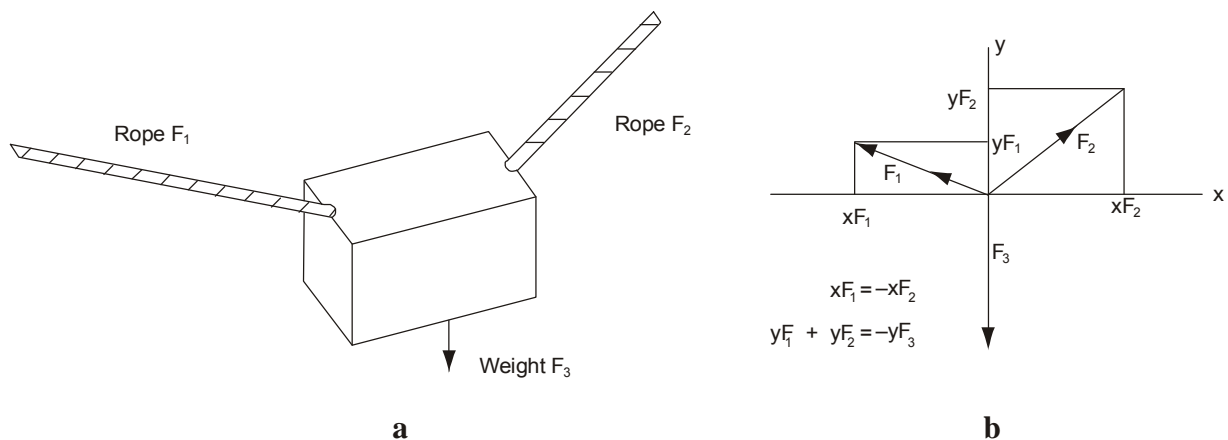
$$\therefore x = 6.69 \text{ metres.}$$

## Equilibrium

10. Two types of equilibrium condition may be recognized, translational equilibrium and rotational equilibrium.

11. **Translational Equilibrium.** An object is said to be in translational equilibrium if it has constant velocity (including velocity equal to zero). In order to achieve this condition the vector sum of all of the forces acting upon the object must be zero. It is usually convenient to consider the components of the forces in three orthogonal directions ( $x$ ,  $y$ , and  $z$  axes). In this case the algebraic sum of the  $x$ ,  $y$  and  $z$  components must each equal zero. As an example consider Fig 3a which shows a uniform rectangular body being supported by two ropes. Three forces are acting on the body; the tension in the ropes,  $F_1$ , and  $F_2$ , and the weight,  $F_3$ . All three forces may be regarded as coplanar and as acting through the centre of gravity and this simplified situation is shown in the vector diagram, Fig 3b. For translational equilibrium the  $x$  component of  $F_1$  must be equal and opposite to the  $x$  component of  $F_2$ ;  $F_3$  has no  $x$  component. This equality can be shown by constructing verticals from the ends of the vectors to the  $x$ -axis. Also, the sum of the  $y$  components of  $F_1$  and  $F_2$  must be equal and opposite to  $F_3$ . The  $y$  components may be determined by constructing horizontals from the ends of the vectors to the  $y$ -axis.

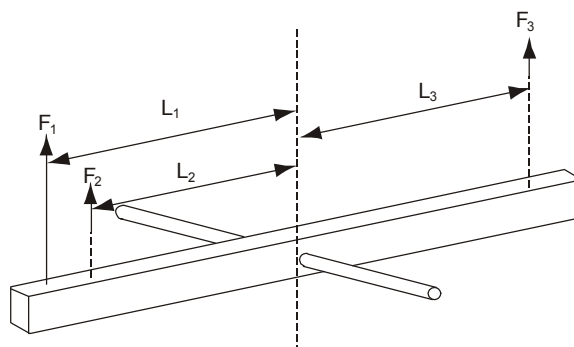
13-29 Fig 3 Translational Equilibrium



12. **Rotational Equilibrium.** An object is in rotational equilibrium when it rotates about an axis of constant direction at a constant angular speed, including zero angular speed. The condition of rotational equilibrium is achieved if the vector sum of all the torques about the axis is zero, i.e. the sum of the clockwise moments has the same magnitude as the sum of the anticlockwise moments.

13. As an example, consider the uniform bar shown in Fig 4, which is free to rotate about the axle. Three forces,  $F_1$ ,  $F_2$ , and  $F_3$ , are shown acting on the bar at distances  $L_1$ ,  $L_2$ , and  $L_3$  respectively from the axle. For rotational equilibrium, the relationship:  $F_1L_1 + F_2L_2 - F_3L_3 = 0$  must be satisfied.

13-29 Fig 4 Rotational Equilibrium



### Stability

14. Three conditions of stability with respect to equilibrium may be recognized as follows:

- Stable Equilibrium.** An object is in stable equilibrium if any small displacement caused by external forces or torques tends to be self-correcting. An example would be a marble in the bottom of a round-bottomed cup. If an external force disturbs its equilibrium it will, after a few oscillations, settle back to its original position.
- Unstable Equilibrium.** An object is in unstable equilibrium if any small displacement tends to be escalating. A marble balanced on the end of a finger would be an example. Any small displacement would cause the marble to move to a totally different position.
- Neutral Equilibrium.** Neutral equilibrium is an intermediate state between stable and unstable equilibrium. Consider a marble at rest on a flat horizontal surface. A small displacement will cause the marble to move but the resulting position is unchanged from an equilibrium point of view from the original position. It will have no tendency either to be displaced further, or to return to its original position.

### Friction

15. When two solid surfaces which are in contact move, or tend to move, relative to each other, a force acts in the plane of contact of the surfaces in a direction opposing the motion. This force is known as friction, and is a result of interactions between the molecules of the two surfaces.

16. **Dynamic Friction.** If an object is sliding over a surface at a constant velocity, then the friction is an example of dynamic friction and there will be some conversion of the kinetic energy of the object into heat. The force of dynamic friction,  $f_d$ , depends upon the material of the two surfaces, on their smoothness and on the component of the force,  $F$ , that presses the two surfaces together. The frictional force is practically independent of the relative velocity of the two surfaces. It has been found that:

$$f_d = u_d F$$

where  $u_d$  is a constant for a given pair of surfaces and is known as the coefficient of dynamic friction. The value of  $u_d$  varies from about 0.06 for a smooth steel surface sliding on ice, to 0.7 for rubber sliding over dry concrete.

17. **Static Friction.** If an object is placed on a plane whose angle of inclination can be increased, it is found that the object remains stationary until a certain angle of inclination is reached whereupon the

object starts to move. The force preventing the object from moving is known as static friction. As the plane's inclination is increased then the value of the component of the weight parallel to the plane increases. As the friction force is equal and opposite to this force (otherwise the object would move), then it too must increase as the inclination is increased until it reaches a critical angle. As with dynamic friction the maximum force of static friction,  $f_{s \max}$  depends upon the materials, the smoothness of the surfaces in contact and on the component of the force,  $F$ , pressing the two surfaces together. For any pair of surfaces it is found that:

$$f_{s \max} = u_s F$$

where  $u_s$  is constant for any two materials of specified smoothness, and is called the coefficient of static friction. It varies from about 0.1 for steel on ice to about 1 for rubber on dry concrete. The coefficient of static friction is always higher than that of dynamic friction, which is why objects tend to start moving with a jerk.

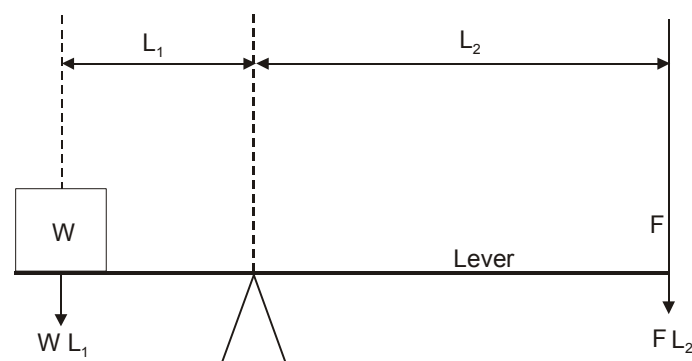
18. **Applicability.** It should be noted that the relationships described above are derived from observations rather than from any theoretical understanding of the mechanisms causing friction. Thus the equations do not have universal applicability; deviations occur, for example, at extreme speeds, when the surfaces in contact are very small and when the force pressing the surfaces together is very large.

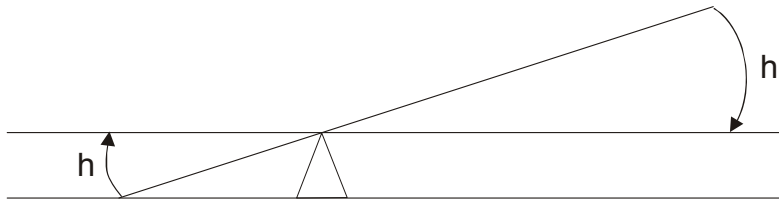
## Machines

19. A machine is any device which enables energy to be used in a convenient way to perform work. Typical examples of simple machines are levers, winches, inclined planes, pulleys and screws. Machines do not save work; in general they allow a smaller force to be applied in order to achieve a result, but the smaller force must be applied over a greater distance. As examples of machines, the lever and a pulley system will be reviewed.

20. **The Lever.** A lever can be described as a rigid beam supported at a point or fulcrum that is fixed, and about which the beam can turn. The arrangement is shown in Fig 5 where the purpose of the lever is to lift a load, of weight  $W$ . A force  $F$  is applied at the opposite end such that the lever is maintained in a horizontal position. In this situation, the system is in rotational equilibrium and so the moments about the pivot must be equal and opposite. The anticlockwise moment is  $WL_1$  whilst the clockwise moment is  $FL_2$ . Therefore, if  $L_2$  is greater than  $L_1$  the force required to balance the load is less than the load. However, as shown in Fig 6, in order to raise the load over a distance,  $h$ , the smaller force must be applied over a greater distance,  $h_1$ .

13-29 Fig 5 Lever in Rotational Equilibrium



**13-29 Fig 6 Lever - Smaller Force over Greater Distance**

21. **A Two-pulley System.** A two-pulley system is shown in Fig 7 in which a force,  $F$ , is applied through a distance,  $a$ , in order to raise a weight,  $W$ , through a distance,  $b$ . From the principle of the conservation of energy (see Volume 13, Chapter 31, Para 16), the potential energy gained by the load will equal the work done by the effort, ignoring incidental energy losses (eg friction). Therefore:

$$Wb = Fa, \text{ or } \frac{W}{F} = \frac{a}{b}$$

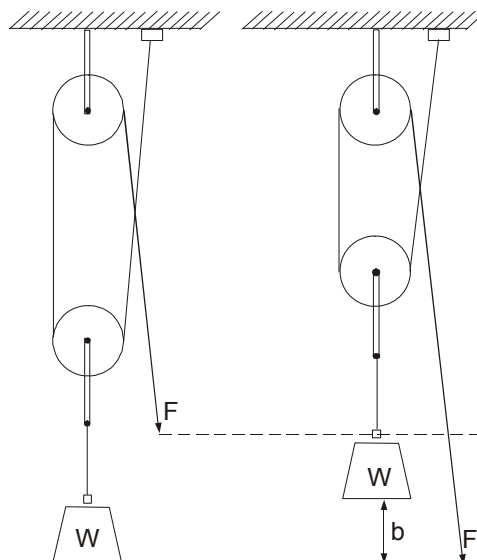
As with the lever, a load can be lifted with a smaller force but that force must be applied over a greater distance.

22. **Mechanical Advantage.** The ratio  $\frac{W}{F}$  is known as the mechanical advantage. For the pulley system shown, if  $a = 2b$ , the mechanical advantage is 2.

23. **Efficiency.** Efficiency is related to the work done in moving an object. Work is described fully in Volume 13, Chapter 31, Para 10 and is the product of a force and the distance moved in the direction of that force. In the example:

$$\text{Efficiency} = \frac{\text{Work done on load}}{\text{Work done by effort}} = \frac{Wb}{Fa}$$

Efficiency is usually expressed as a percentage. Since friction has to be overcome, the efficiency of a machine is always less than 100%.

**13-29 Fig 7 A Two-pulley System**

## CHAPTER 30 - KINEMATICS

Contents	Page
Introduction .....	1
LINEAR MOTION .....	1
Speed and Velocity.....	1
Acceleration .....	1
Vector Representation.....	2
Relationship between Distance, Velocity, Acceleration, and Time .....	2
Relative Velocity .....	3
ROTARY MOTION .....	4
Angular Velocity.....	4
Relationship between Angular and Linear Velocity.....	5
Vector Representation of Angular Quantities.....	6

### Table of Figures

13-30 Fig 1 Velocity - Time Graph .....	3
13-30 Fig 2 Relative Velocity.....	4
13-30 Fig 3 The Acceleration Experienced in Uniform Circular Motion .....	4
13-30 Fig 4 Vector Representation of Angular Quantities .....	6

### Introduction

1. Kinematics is the study of motion without reference to the forces involved. In this chapter, linear and angular motion will be examined.

## LINEAR MOTION

### Speed and Velocity

2. Speed is the ratio of the distance covered by a moving body, in a straight line or in a continuous curve, to the time taken. The velocity of a body is defined as its rate of change of position with respect to time, the direction of motion being specified. If the body is travelling in a straight line, it is in linear motion, and if it covers equal distances in equal successive time intervals it is in uniform linear motion.

3. For uniform velocity, where  $s$  is the distance covered in time  $t$ , the velocity  $v$  is given by:

$$v = \frac{s}{t}$$

4. In the more general case, the instantaneous velocity  $v_i$  is given by:

$$v_i = \frac{ds}{dt}$$

### Acceleration

5. The acceleration of a body is its rate of change of velocity with respect to time. Any change of either speed or direction of motion involves an acceleration; a retardation is merely a negative acceleration.

6. When the velocity of a body changes by equal amounts in equal intervals of time it is said to have a uniform acceleration, measured by the change in velocity in unit time.

7. If the initial velocity  $u$  of a body in linear motion changes uniformly in time  $t$  to velocity  $v$ , its acceleration  $a$  is given by:

$$a = \frac{(v - u)}{t}$$

### Vector Representation

8. Velocity and acceleration are vector quantities and the laws of vector addition may be applied. Thus, the resultant velocity of a body having two separate velocities (e.g. an aircraft flying in wind) may be found by the parallelogram law. In addition, a single velocity may be resolved into two or more components.

### Relationship between Distance, Velocity, Acceleration, and Time

9. Some useful formulae can be derived relating distance, velocity, acceleration, and time. When dealing only with motion in a straight line, the directional aspects of velocity and acceleration can be ignored, apart from the use of positive and negative signs to indicate forward and backward motion.

10. Consider a body with initial velocity,  $u$ , which in time  $t$  attains, under uniform acceleration  $a$ , a final velocity  $v$ . Suppose that the distance covered during this time is  $s$ . It has been stated that:

$$a = \frac{(v - u)}{t}$$

$$\therefore at = v - u, \text{ or}$$

$$v = u + at \dots \dots \dots (1)$$

The mean or average velocity of the body is  $(u + v)/2$ . Therefore, the distance covered in time  $t$  will be given by:

$$s = \frac{(u + v)}{2} \cdot t \dots \dots \dots (2)$$

Substituting for  $v$  from equation (1),

$$s = \frac{1}{2} (u + u + at) \cdot t, \text{ or}$$

$$s = ut + \frac{1}{2} at^2 \dots \dots \dots (3)$$

From equation (1),  $t = (v - u)/a$ . Substituting for  $t$  in equation (2),

$$s = \frac{1}{2} \frac{(u + v)(v - u)}{a}, \text{ or}$$

$$v^2 = u^2 + 2as \dots \dots \dots (4)$$

11. Note that if the distance travelled in time  $t$  is denoted by  $s$ , velocity can be obtained by differentiation:

$$v = \frac{ds}{dt}$$

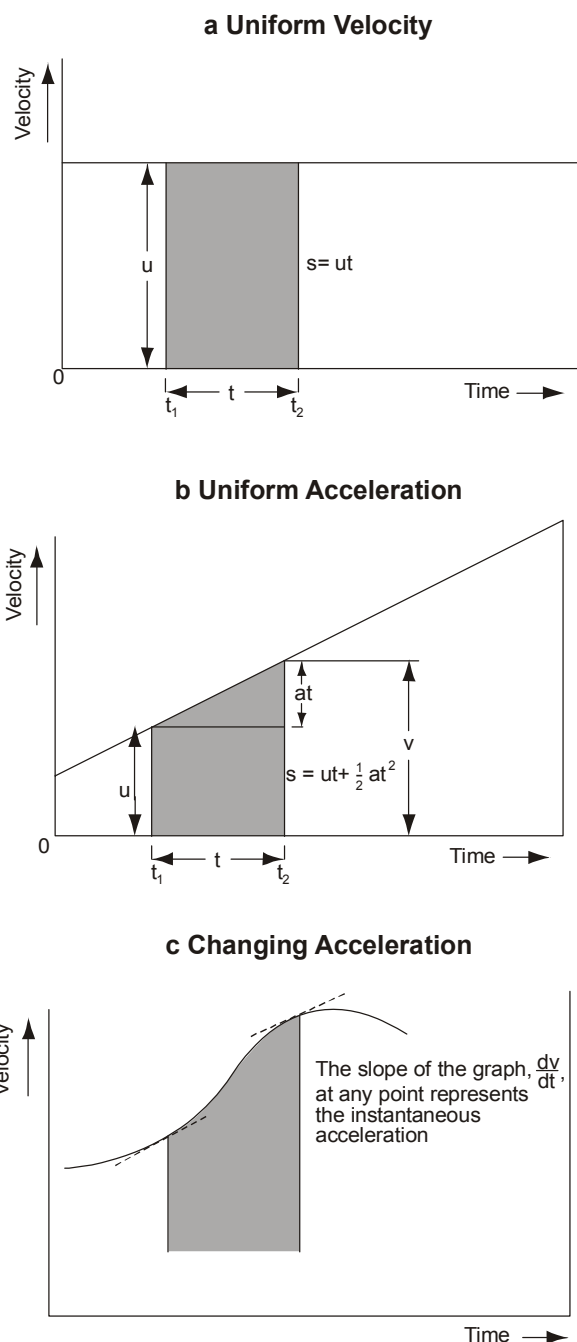
Similarly, the acceleration  $a$  is given by:

$$a = \frac{dv}{dt} = \frac{d^2s}{dt^2}$$

Conversely, given an expression for acceleration, integration will give an expression for velocity, and further integration an expression for displacement.

12. *Velocity-Time Graphs.* The linear motion of a body can be illustrated by means of velocity-time graphs, examples of which are shown in Fig 1. In each case the distance travelled by the body between times  $t_1$  and  $t_2$  is represented by the area under the corresponding part of the graph (shaded in Fig 1). The instantaneous acceleration at any time is given by the slope of the curve at that point.

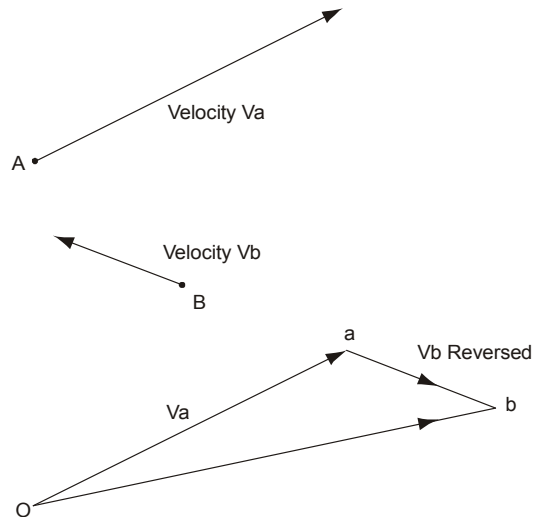
**13-30 Fig 1 Velocity - Time Graph**



### Relative Velocity

13. It is sometimes necessary to determine the velocity with which one moving body appears to be moving with respect to another. This is known as the relative velocity. For linear motion this type of problem may be solved graphically, as in Fig 2, by drawing from an origin a vector representing the velocity of body A, and from the end of this vector drawing a vector to represent the velocity of B reversed. In Fig 2 the third side of the triangle, ob, represents the velocity of A relative to B.

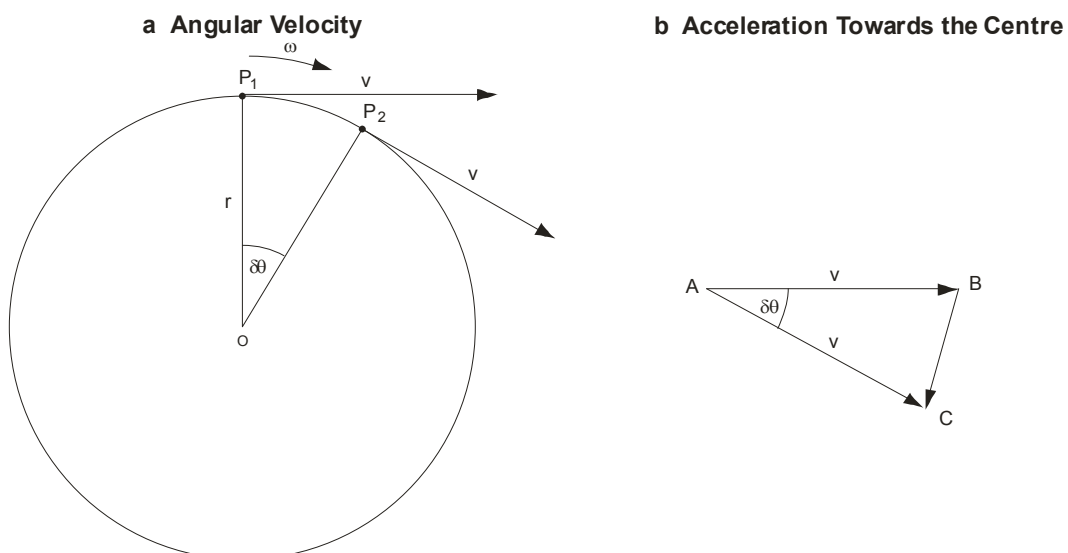


**13-30 Fig 2 Relative Velocity****ROTARY MOTION****Angular Velocity**

14. So far, only motion in a straight line has been examined. Consider now a point P which moves in a circle of radius  $r$  at constant speed  $v$ . Its angular velocity  $\omega$  is given by  $d\theta/dt$  where  $\theta$  is the angular displacement in radians (Fig 3a). The speed  $v$  is given by:

$$v = r \cdot \frac{d\theta}{dt} = r\omega$$

15. Although the velocity of the point is constant in magnitude, it is constantly changing in direction, and therefore, by definition, P is subject to an acceleration. Consider the point P at a certain instant (position  $P_1$  in Fig 3a), and also after a small interval of time  $\delta t$  (position  $P_2$ ). It is clear that the vectors representing motion at these two instants are of the same length, but are in different directions. They are represented by AB and AC in Fig 3b, BC representing the change in velocity in the time  $\delta t$ . The acceleration is given by  $BC/\delta t$ .

**13-30 Fig 3 The Acceleration Experienced in Uniform Circular Motion**

16. Since  $\delta t$  is small, the angular displacement,  $\delta\theta$ , is also small, and BC may be considered to be almost perpendicular to both AB and AC, that is, directed towards the centre of the circle.

$$BC = v \cdot \delta\theta$$

$$\text{but, } \delta\theta = \frac{P_1P_2}{r}$$

If the time interval  $\delta t$  is very small, the straight line  $P_1P_2$  is very nearly equal to the distance  $P_1P_2$  along the arc, which is  $v \cdot \delta t$ .

$$\therefore \delta\theta = \frac{v \cdot \delta t}{r} \text{ radians}$$

$$\therefore BC = \frac{v^2 \delta t}{r}$$

$$\text{Acceleration} = \frac{BC}{\delta t}$$

$$= \frac{v^2}{r} \text{ or } v\omega, \text{ or } \omega^2 r$$

all towards the centre.

17. To summarize, a body moving at constant speed  $v$  in a circle of radius  $r$  has a constant acceleration of  $v^2/r$ , directed towards the centre of the circle.

18. The following formulae for circular motion, similar to those for linear motion, may be derived:

For uniform angular velocity,  $\theta = \omega t$ .

For uniform angular acceleration,

$$\omega_2 = \omega_1 + \alpha t \quad (\text{cf, } v = u + at)$$

$$\theta = \frac{(\omega_1 + \omega_2)t}{2}$$

$$\text{and } \theta = \omega_1 t + \frac{1}{2} \alpha t^2 \quad (\text{cf, } s = ut + \frac{1}{2} at^2)$$

$$\text{and } \omega^2 = \omega_1^2 + 2\alpha\theta \quad (\text{cf, } v^2 = u^2 + 2as)$$

where

$\omega_1$  = initial velocity in radians per sec

$\omega_2$  = velocity in radians per sec after  $t$  sec

$\alpha$  = angular acceleration in radians per sec<sup>2</sup>

$\theta$  = angle through which turned, in radians

### Relationship between Angular and Linear Velocity

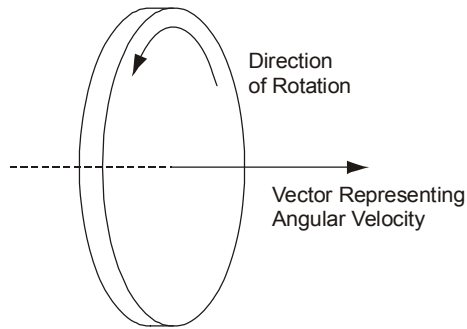
19. Consider again the point P, rotating with uniform angular velocity  $\omega$  and radius  $r$  (Fig 3a). The length of the arc  $P_1P_2$  is the radius multiplied by the angle in radians, i.e. arc  $P_1P_2 = r\theta$ . But, for uniform angular velocity,  $\theta = \omega t$ , hence arc  $P_1P_2 = r\omega t$  and the linear velocity of P is  $r\omega t/t$ , or  $\omega r$ . Similarly, the linear acceleration of P is equal to the angular acceleration times the radius; its direction is radially towards the centre of rotation. Summarizing, if  $v$  = linear velocity,  $\omega$  = angular velocity,  $a$  = linear acceleration, and  $\alpha$  = angular acceleration, then:

$$v = \omega r, \text{ or } \omega = \frac{v}{r} \text{ and } a = \alpha r, \text{ or } \alpha = \frac{a}{r}$$

**Vector Representation of Angular Quantities**

20. Angular velocity and acceleration are vector quantities. By convention, they are represented by vectors; the vector length represents the magnitude of the quantity and its direction is perpendicular to the plane of rotation, i.e. parallel to the axis of rotation. The direction of the arrow is such that, on looking in its direction, the rotation is clockwise (Fig 4). This convention is known as the right-handed screw law. Such vectors can be combined and resolved according to the normal principles of vectors.

**13-30 Fig 4 Vector Representation of Angular Quantities**



## CHAPTER 31 - DYNAMICS

Contents	Page
Introduction .....	1
QUANTITIES .....	1
Mass .....	1
Density .....	2
Momentum .....	2
Force .....	2
Conservation of Momentum .....	2
Work .....	2
Energy .....	3
Power .....	3
CIRCULAR MOTION .....	4
Centripetal Force .....	4
Moment of Inertia .....	4
SIMPLE HARMONIC MOTION .....	4
General .....	4
The .....	5
Simple Pendulum .....	5
The Compound Pendulum .....	7
A Mass Supported by a Spring .....	7

### Figures

13-31 Fig 1 Simple Harmonic Motion .....	5
13-31 Fig 2 The Simple Pendulum .....	6

### Introduction

1. Dynamics is the study of motion related to force. In Volume 13, Chapter 30, it was shown how a body moves; in this chapter it will be shown why the body moves in that way.

## QUANTITIES

### Mass

2. When an object is at rest, a force is necessary to make it move; similarly, a body in motion needs a force to be applied in order to change its motion. This reluctance to any change in motion is called inertia and the property that gives rise to inertia is mass. It can be shown that inertia is directly proportional to mass.

3. The mass of a body may be defined as the quantity of matter in the body. The unit of mass is the kilogram (kg) and the standard kilogram is a cylinder of platinum-iridium alloy kept at the Bureau International des Poids et Mesures.

## Density

4. The density of a substance is the mass per unit volume of that substance. The unit of density is the kilogram per cubic metre ( $\text{kg m}^{-3}$ ). Relative density (or specific gravity) is the ratio of the density of a substance at a stated temperature to the density of water at 4 °C.

## Momentum

5. The momentum of a body is defined as the product of its mass and its velocity. As the definition includes the velocity term, momentum is a vector quantity and a change in either speed or direction constitutes a change in momentum. The unit of momentum is the kilogram metre per second ( $\text{kg ms}^{-1}$ ).

## Force

### 6. Newton's laws of motion:

- a. **First Law.** A body remains in a state of rest or uniform motion (i.e. no acceleration) in a straight line unless acted upon by an external force.
- b. **Second Law.** The rate of change of momentum is proportional to the applied force, and the change of momentum takes place in the direction of the applied force.
- c. **Third Law.** Every action is opposed by an equal and opposite reaction.

7. By selecting a unit of force as that force which gives unit acceleration to unit mass the second law may be written as:

$$F = ma$$

The unit of force is the newton (N) which is the force necessary to induce an acceleration of  $1\text{ms}^{-2}$  in a mass of 1 kilogram.

## Conservation of Momentum

8. Consider two bodies, A and B, travelling in the same direction, which collide, the duration of the collision being the short time  $t$ . Throughout the collision, each will experience a force equal and opposite to that experienced by the other (Newton's third law). The impulse of the force is  $Ft$ , and is the same for each body. Thus, the change of momentum will be the same for each body. If at the time of collision body A was overtaking body B, it is apparent that the effect of the impact will be to decrease the momentum of A and increase that of B, and the total momentum of the system of two bodies will be unchanged.

9. The Law of Conservation of Momentum. The effects of the interaction of parts of a closed system are summarized in the Law of Conservation of Momentum, which states that the total momentum in any given direction before impact is equal to the total momentum in that direction after impact.

## Work

10. A force is said to do work when its point of application moves, and the amount of work done is the product of the force and the distance moved in the direction of the force.

11. The unit of work is the joule (J) which is the work done when a force of 1 newton moves 1 metre in the direction of the force.

12. The gravitational force acting on a body is the product of its mass ( $m$ ) and the acceleration of gravity ( $g$ ), i.e.  $mg$ . Therefore, the work done against gravity in raising the body through a vertical distance,  $h$ , is  $mgh$ .

### Energy

13. The energy possessed by a body is its capacity to do work; the unit of energy is the joule. A body may possess this capacity by virtue of:

- a. Its position, when the energy is called potential energy (PE).
- b. Its motion, when the energy is called kinetic energy (KE).

14. Consider a mass,  $m$ , projected vertically upwards from the ground with initial velocity,  $u$ . It is acted upon (downwards) by gravity, and will attain a height,  $h$ , which can be determined by substitution in the formula  $v^2 = u^2 + 2as$ , thus:

$$0 = u^2 + 2(-g)h$$

the velocity being zero at the highest point

$$\therefore h = u^2/2g$$

The work done reaching a height  $h$  is  $mgh$ . Substituting for  $h$ ,

$$\text{work done} = mgu^2/2g = \frac{1}{2}mu^2$$

$$= \text{work done in coming to rest from velocity } u$$

$$= \text{KE.}$$

15. At the highest point, where the velocity is zero, the kinetic energy is zero. If, however, the mass is allowed to fall, it will, at the moment of striking the ground, have regained its original velocity, and thus its original kinetic energy. The energy which it possessed at the highest point is potential energy and equals its original kinetic energy.

16. Consider also a body acted upon by a constant force  $F$ .

$$\text{Then } F = ma = m \left( \frac{v^2 - u^2}{2s} \right)$$

where  $u$ ,  $v$ , and  $s$  have the conventional significance.

$$\text{Thus: } Fs = \frac{1}{2}m(v^2 - u^2)$$

The left-hand side is the work done by the force, while the right-hand side is the resulting change in kinetic energy. The general expression for the kinetic energy of a body moving with velocity  $v$  is:

$$\text{KE} = \frac{1}{2}mv^2$$

This illustrates the principal of the conservation of energy - energy can be neither created nor destroyed, though it may be converted from one form to another. It should be noted that kinetic and potential energies are examples of one form of energy, namely mechanical energy; a change may involve other forms of energy such as chemical, heat, light, sound, magnetic or electrical energy.

### Power

17. Power is the rate of doing work and has the unit, watt (W), equal to 1 joule per second.

## CIRCULAR MOTION

### Centripetal Force

18. It was shown in Volume 13, Chapter 30 that a body travelling with uniform speed in a circle has an acceleration of  $v^2/r$  towards the centre of the circle. The force producing this acceleration is termed centripetal force, and for a body of mass  $m$  the centripetal force is  $mv^2/r$  towards the centre of the circle. It should be noted that in the case of a body travelling in a circular path at the end of a string, while the mass is experiencing centripetal force towards the hand, there is an equal and opposite reaction on the hand holding the string, known as centrifugal force. Centrifugal force exists only as an equal and opposite reaction to centripetal force.

### Moment of Inertia

19. Consider now a rigid body, such as a flywheel, free to rotate about an axis through its centre, at angular velocity  $\omega$ . When the wheel is rotating all the particles of the wheel have the same angular velocity, but their linear velocities will depend on their individual distances from the axis, those on the rim moving much faster than those near the axis.

20. A particle of mass  $m$ , distance  $r$  from the centre, and with linear velocity  $v$ , has kinetic energy  $\frac{1}{2}mv^2$  or  $\frac{1}{2}mr^2\omega^2$ . The total kinetic energy of all the particles is  $\frac{1}{2}\omega^2 \sum mr^2$ .

21. The sum (expressed as  $\Sigma$ ) of the products  $mr^2$  for all the particles in a rigid body rotating about a given axis is called the Moment of Inertia ( $I$ ).

$$I = \Sigma mr^2$$

22. Moment of inertia can be considered as the rotational equivalent of mass and, just as in linear motion, if a force is applied to a body, force = mass  $\times$  acceleration, so in circular motion, if a torque is applied to a wheel,

$$\text{torque} = \text{moment of inertia} \times \text{angular acceleration}.$$

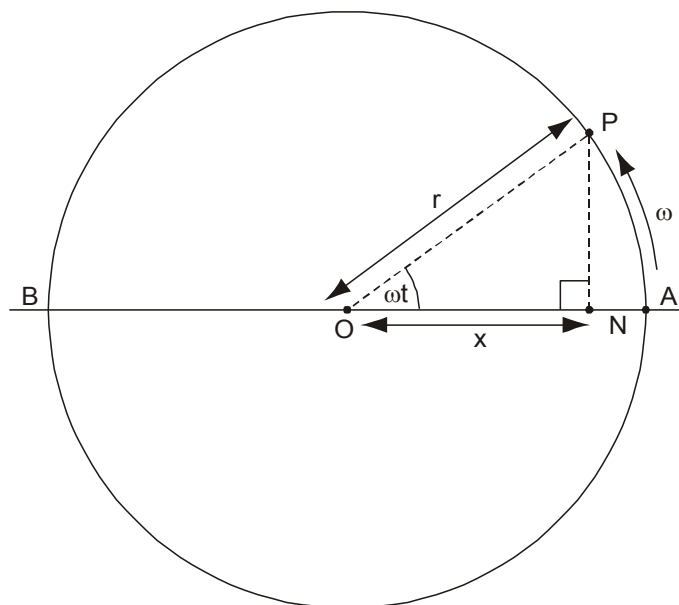
23. **Radius of Gyration.** The radius of gyration is a useful concept whereby the mass  $M$  ( $=\Sigma m$ ) of the wheel is considered to be concentrated in a ring of radius  $k$  from the axis, such that  $\Sigma mr^2 = Mk^2$ . 'k' is then known as the radius of gyration, and  $I = Mk^2$ .

## SIMPLE HARMONIC MOTION

### General

24. A type of motion which is frequently encountered is simple harmonic motion. This is defined as the motion of a body which moves in a straight line so that its acceleration is directly proportional to the distance from a fixed point in the line, and always directed towards that point. A mass suspended from a spiral spring which is given a small vertical displacement from its equilibrium position and then released will oscillate with simple harmonic motion.

25. Simple harmonic motion can be illustrated by considering the projection of a point moving at uniform speed in a circular path on to a diameter of that circle (Fig 1).

**13-31 Fig 1 Simple Harmonic Motion**

Let the point P start at position A, and move with constant angular velocity  $\omega$ , in a circle of radius  $r$ . Let PN be the perpendicular from the point to the diameter AOB. Then while the point P moves round the circle from A through B back to A, N moves through O to B and back to A. The time for one complete cycle of oscillation is  $2\pi/\omega$  and this is called the periodic time or simply the period of the oscillation. The frequency of oscillation is the number of complete cycles in unit time, and frequency =  $1/\text{periodic time} = \omega/2\pi$ . The distance ON at any time,  $t$ , is given by  $x = r \cos \omega t$ . The instantaneous velocity of N is given by:

$$\frac{dx}{dt} = -r\omega \sin \omega t \text{ and the acceleration by:}$$

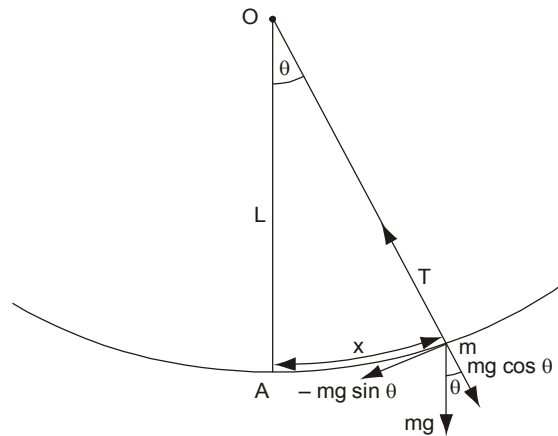
$$\frac{d^2x}{dt^2} = -r\omega^2 \cos \omega t = -\omega^2 x$$

Since the angular velocity  $\omega$  is constant, the acceleration is proportional to the distance ON, thus the movement of N accords with the definition of simple harmonic motion. Note that the expression is negative, since the acceleration is measured in the opposite direction to that of the displacement ON (i.e. towards O).

**The Simple Pendulum**

26. The simple pendulum provides an example of simple harmonic motion, provided that its amplitude of oscillation is small. Consider a mass,  $m$ , suspended from a light inextensible cord of length  $L$ , displaced through a small angle  $\theta$  (in radians) from the vertical (Fig 2).



**13-31 Fig 2 The Simple Pendulum**

27. The forces acting on the mass are its weight,  $mg$ , acting vertically downwards, and the tension,  $T$ , in the cord. The weight can be resolved into two components,  $mg \cos \theta$  equal and opposite to  $T$ , and the restoring force,  $-mg \sin \theta$  acting in the opposite direction to the direction of displacement of the mass from its central position. Using the equation:

force = mass  $\times$  acceleration,

$$-mg \sin \theta = ma$$

$$\therefore a = -g \sin \theta, \text{ or, for small angles, } a = -g\theta \text{ (approximately)}$$

$$\text{but, } \theta = \frac{x}{L}$$

$$\therefore a = -\frac{gx}{L}$$

28. As  $g$  and  $L$  are constant for a particular location and a particular pendulum, the acceleration is proportional to the displacement  $x$ , and acts in the direction of the force towards the mid point  $A$ , thus (for small displacements) satisfying the conditions for simple harmonic motion.

29. By comparing the results of paras 25 and 27 it can be seen that the period of oscillation of a simple pendulum, for small displacements, is given by:

$$\text{period} = 2\pi \sqrt{\frac{L}{g}} \text{ sec,}$$

and the frequency by:

$$\text{frequency} = \frac{1}{2\pi} \sqrt{\frac{g}{L}} \text{ hertz.}$$

It will be noted that, for small angles of swing, the period of oscillation of a simple pendulum is independent of the mass, and of the amplitude of swing.

**The Compound Pendulum**

30. For a compound, or rigid, pendulum, the period is given by:

$$\text{period} = 2\pi\sqrt{\frac{I}{mgh}}$$

where  $I$  is the moment of inertia about the axis of rotation, and  $h$  is the distance from the centre of gravity to the pivot.

**A Mass Supported by a Spring**

31. The comparable equation for the period of oscillation of a mass  $m$  kilograms suspended from a spring of stiffness  $e$  kilograms per metre is:

$$\text{period} = 2\pi\sqrt{\frac{m}{e}} \text{ sec.}$$

## CHAPTER 32 - HYDRAULICS

Contents	Page
Introduction .....	1
Definition of Terms .....	1
Transmission of Force and Motion by Fluids .....	1
Components of a Typical Hydraulic System .....	3

### Table of Figures

13-32 Fig 1 A Simple Hydraulic System.....	2
13-32 Fig 2 Relative Movement in a Simple Hydraulic System .....	2
13-32 Fig 3 Typical Practical Hydraulic System.....	3

### Introduction

1. Hydraulic systems provide a means of transmitting a force by the use of fluids. They are concerned with the generation, modulation and control of pressure and flow of the fluid to provide a convenient means of transmitting power for the operation of a wide range of aircraft services. A typical aircraft hydraulic system will be used for operating flying controls, flaps, retractable undercarriages and wheelbrakes. Hydraulic systems can transmit high forces with rapid, accurate response to control demands.

### Definition of Terms

2. The following definitions need to be understood:

a. **Pressure.** Pressure is the force per unit area exerted by a fluid on the surface of a container. Pressure is measured in bars, pascals (Pa), or newtons per square metre (N/m<sup>2</sup>).

$$1 \text{ bar} = 100,000 \text{ Pa} = 100,000 \text{ N/m}^2$$

b. **Force.** The force exerted on a particular surface by a pressure is calculated from the formula:

$$\text{Force} = \text{Pressure} \times \text{Surface Area.}$$

c. **Fluid.** A fluid is a liquid or gas which changes its shape to conform to the vessel that contains it.

d. **Hydraulic Fluid.** Hydraulic fluid is an incompressible oil. In aircraft systems, low flammability oils are used, the boiling and freezing points of which fall outside operating parameters.

### Transmission of Force and Motion by Fluids

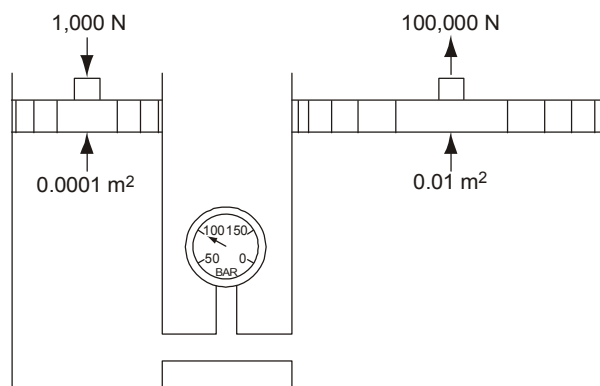
3. When a force is applied to one end of a column of confined fluid a pressure is generated which is transmitted through the column equally in every direction. Fig 1 illustrates a simple hydraulic system. Two

cylinders are connected by a tube. The cylinders contain pistons of surface areas  $0.01 \text{ m}^2$  and  $0.0001 \text{ m}^2$  giving a piston area ratio of 100:1. The system is filled with hydraulic fluid and fitted with a pressure gauge. A force of 1,000 newtons (N) is applied to the small piston. The force will produce a pressure (P) in the fluid so:

$$P = \frac{\text{Force (F)}}{\text{Piston Area (A)}} \quad \text{i.e.} \quad P = \frac{1000}{0.0001} = 10,000,000 \text{ Pa or } 10 \text{ MPa, which is } 100 \text{ bar.}$$

The system pressure will appear on the gauge and will be felt on every surface within the system. Thus at the large piston a force (F) will be exerted where  $F = P \times A = 10 \times 10^6 \times 0.01 = 100,000 \text{ N}$ . The force applied at the small piston is therefore increased on delivery by the large piston by a factor of 100, i.e. in the same ratio as that of piston area. This is sometimes referred to as force multiplication.

**13-32 Fig 1 A Simple Hydraulic System**



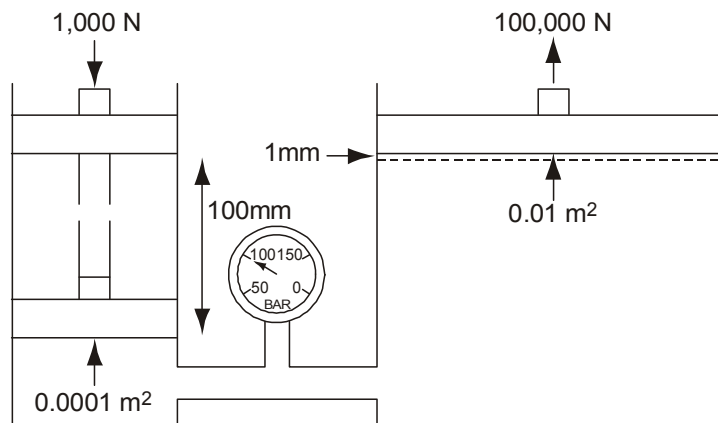
4. If the small piston is now moved down in its cylinder through a distance of 100 mm, as illustrated in Fig 2, the large piston will move upwards through a distance inversely proportional to the piston area ratio, ie through 1 mm. The work done by the small force is transmitted hydraulically and equals the work expended in moving the greater force through a smaller distance,

$$\text{i.e.} \quad \text{force 1} \times \text{distance 1} = \text{force 2} \times \text{distance 2}$$

$$\text{or} \quad 1000 \text{ N} \times 100 \text{ mm} = 100,000 \text{ N} \times 1 \text{ mm}$$

This is the principle of the hydraulic lever and is the operating principle of any hydraulic system.

**13-32 Fig 2 Relative Movement in a Simple Hydraulic System**

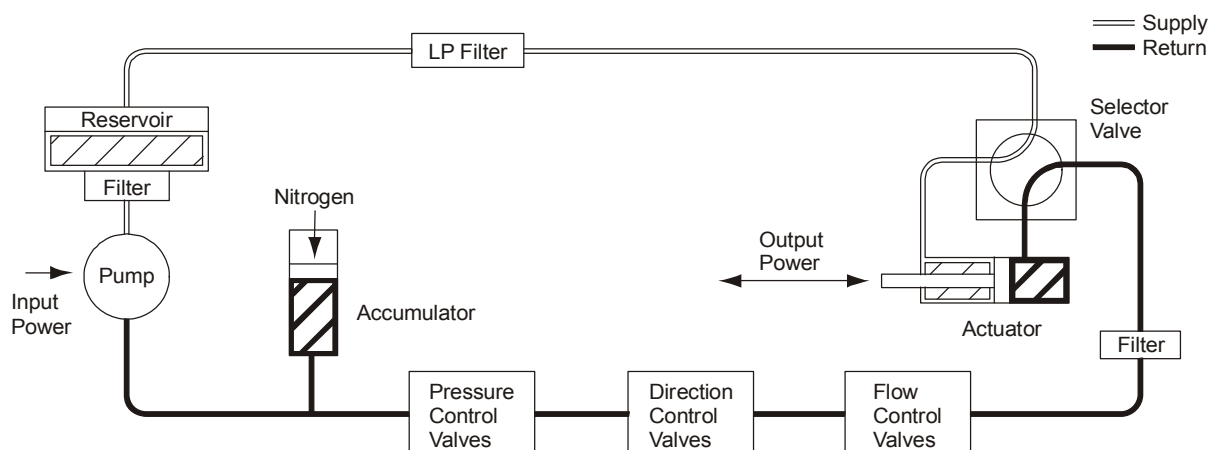


## Components of a Typical Hydraulic System

5. In a simple hydraulic system, force applied by hand or foot to a small 'master' piston is transmitted to a larger 'slave' piston to operate the required service. In more complex, high performance systems, the master piston is replaced by hydraulic pumps and the slave by actuators driving each of the powered services. A typical system illustrated in Fig 3 will operate at between 200 and 300 bar, and it will include the following components:

- a. **Hydraulic Pump.** The pump generates hydraulic pressure and delivers it to the pressure lines in the system. It will usually be either engine driven or electrically powered.
- b. **Valves.** Non-return valves control the direction of fluid flow, pressure relief valves the level of power produced, and selector valves the amount of fluid flow to related actuators.
- c. **Actuators.** Actuators convert the hydraulic power into usable mechanical power at the point required.
- d. **Hydraulic Fluid.** The fluid provides the means of energy transmission as well as lubrication, and cooling of the system.
- e. **Connectors.** The connectors link the various system components. They are usually rigid pipes, but flexible hoses are also used.
- f. **Reservoir.** Fluid is stored in a system reservoir in sufficient quantity and quality to satisfy system requirements. The fluid becomes heated by operation of the system, and the reservoir performs the secondary functions of cooling the fluid and of allowing any air absorbed in the fluid to separate out.
- g. **Filters.** Hydraulic system components are readily damaged by solid particles carried in the fluid, and several stages of filtration are included in a system to prevent debris passing from one component to the next. The filters perform as useful tell-tales of fluid contamination. In addition, samples are taken periodically from the fluid and analysed to detect trends in acid and other trace element levels.
- h. **Accumulator.** An accumulator is a cylinder containing a floating piston. On one side of the piston is nitrogen at system pressure, and on the other hydraulic fluid from the pressure line. When the hydraulic pressure is increased, the nitrogen is compressed. The compressed nitrogen then acts as a spring and can damp out system pressure ripples. It also acts as a reserve of fluid and an emergency power source.

13-32 Fig 3 Typical Practical Hydraulic System



## CHAPTER 33 - INTRODUCTION TO GYROSCOPES

Contents	Page
Introduction .....	2
Definition of Terms .....	4
Classification of Gyroscopes .....	5
LAWS OF GYRODYNAMICS .....	5
Rigidity in Space .....	5
The First Law of Gyrodynamics .....	5
Precession .....	6
The Second Law of Gyrodynamics .....	6
Direction of Precession .....	7
CONSERVATION OF ANGULAR MOMENTUM .....	7
Explanation .....	7
Cause of Precession .....	8
Gyroscopic Resistance .....	8
Secondary Precession .....	10
THE RATE GYROSCOPE .....	10
Principle of Operation .....	10
THE RATE-INTEGRATING GYROSCOPE .....	12
Principle of Operation .....	12
THE DISPLACEMENT GYROSCOPE .....	14
Definition .....	14
Wander .....	15
Earth Rotation .....	15
Transport Wander .....	16
Apparent Wander Table .....	18
Practical Corrections for Topple and Drift .....	19
Gimbal Lock .....	20
Gimbal Error .....	20
OTHER TYPES OF GYRO .....	20
The Ring Laser Gyro .....	20
Fibre Optic Gyros .....	23
Vibrating Gyros .....	23

## Table of Figures

13-33 Fig 1 One-degree-of-freedom Gyroscope .....	3
13-33 Fig 2 Two-degrees-of-freedom Gyroscope .....	3
13-33 Fig 3 Precession.....	6
13-33 Fig 4 Determining Precession .....	7
13-33 Fig 5 Instant of Spin and Precession .....	9
13-33 Fig 6 Gyroscopic Resistance .....	9
13-33 Fig 7 Precession Opposed by Secondary Precession.....	10
13-33 Fig 8 Gyro with One degree of Freedom – Precession.....	11
13-33 Fig 9 Rate Gyroscope .....	11
13-33 Fig 10 Rate of Turn Indicator .....	12
13-33 Fig 11 Simple Rate-integrating Gyroscope .....	13
13-33 Fig 12 Function of Rate-integrating Gyroscope .....	14
13-33 Fig 13 Rate-integrating Azimuth Gyroscope .....	14
13-33 Fig 14 Components of Earth Rate.....	16
13-33 Fig 15 Apparent Drift .....	17
13-33 Fig 16 Transport Wander .....	18
13-33 Fig 17 Gravity Sensitive Switch.....	19
13-33 Fig 18 Gravity Levelling.....	19
13-33 Fig 19 Synchro Case Levelling Device .....	20
13-33 Fig 20 Schematic Diagram of a Ring Laser Gyro .....	21
13-33 Fig 21 Fibre Optic Gyro .....	23
13-33 Fig 22 The 'GyroChip' Vibrating Gyro .....	23

## Tables

Table 1 Classification of Gyros.....	5
Table 2 Types of Wander .....	15
Table 3 Components of Drift and Topple – Earth Rate and Transport Wander Rate .....	18

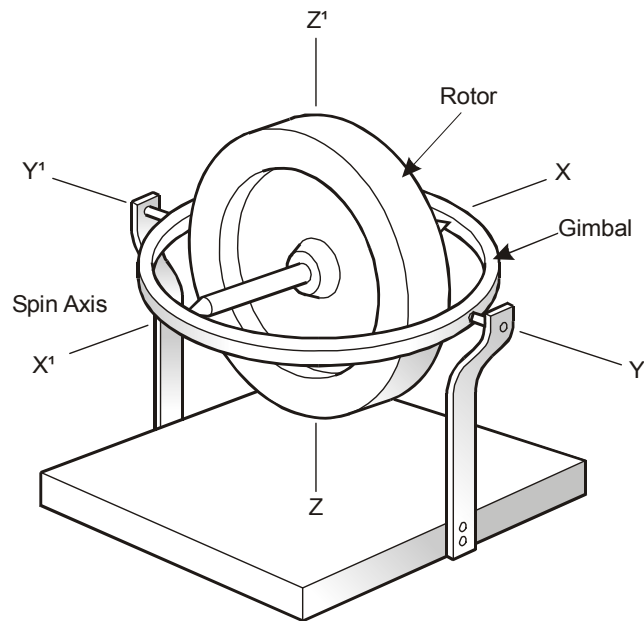
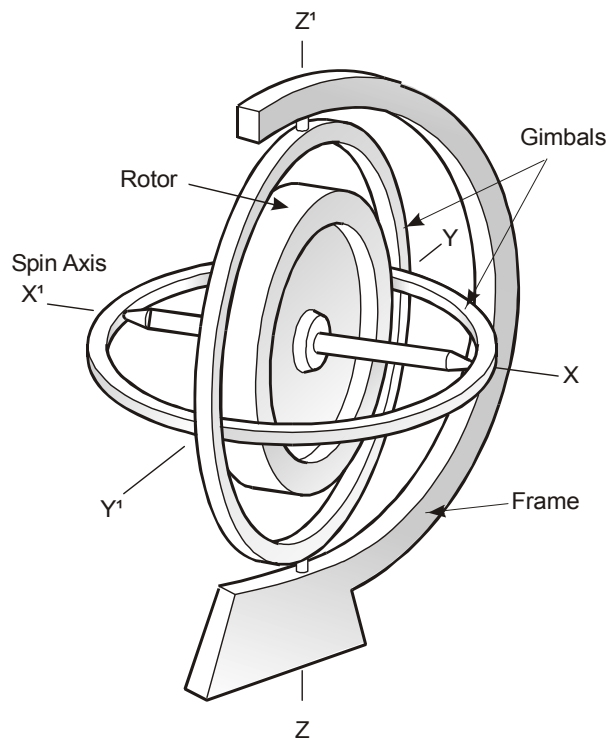
## Introduction

1. Modern technology has brought about many changes to the gyroscope. The conventional spinning gyroscope is still in current use for flight instruments in smaller and simpler aircraft. More sophisticated aircraft however, make use of devices which are termed 'gyros', but this is because of the tasks they perform rather than their manner of operation. Gyroscopes can therefore be categorised as:

- a. Spinning Gyroscopes.
- b. Optical Gyroscopes.
- c. Vibrating Gyroscopes.

This chapter will concentrate for the most part on the spinning gyroscope.

2. A conventional gyroscope consists of a symmetrical rotor spinning rapidly about its axis and free to rotate about one or more perpendicular axis. Freedom of movement about one axis is usually achieved by mounting the rotor in a gimbal, as in Fig 1 where the gyro is free to rotate about the YY<sup>1</sup> axis. Complete freedom can be approached by using two gimbals, as illustrated in Fig 2.

**13-33 Fig 1 One-degree-of-freedom Gyroscope****13-33 Fig 2 Two-degrees-of-freedom Gyroscope**

3. The physical laws which govern the behaviour of a conventional gyroscope are identical to those which account for the behaviour of the Earth itself. The two principal properties of a gyro are rigidity in inertial space and precession. These properties, which are explained later, are exploited in some heading reference and inertial navigation systems (INS) and other aircraft instruments which are described in Volume 5, Chapter 10.



## Definition of Terms

4. The following fundamental mechanical definitions provide the basis of the laws of gyro dynamics:

- a. **Momentum.** Momentum is the product of mass and velocity ( $mv$ ).
- b. **Angular Velocity.** Angular velocity ( $\omega$ ) is the tangential velocity ( $v$ ) at the periphery of a circle, divided by the radius of the circle ( $r$ ), so  $\omega = \frac{v}{r}$ . Angular velocity is normally measured in radians per second.
- c. **Moment of Inertia.** Since a rotating rigid body consists of mass in motion, it possesses kinetic energy. This kinetic energy can be expressed in terms of the body's angular velocity and a quantity called 'Moment of Inertia'. Imagine the body as being made up of an infinite number of particles, with masses  $m_1, m_2$ , etc, at distances  $r_1, r_2$ , etc from the axis of rotation. In general, the mass of a typical particle is  $m_x$  and its distance from the axis of rotation is  $r_x$ . Since the particles do not necessarily lie in the same plane,  $r_x$  is specified as the perpendicular distance from the particle to the axis. The total kinetic energy of the body is the sum of the kinetic energy of all its particles:

$$K = \frac{1}{2} m_1 r_1^2 \omega^2 + \frac{1}{2} m_2 r_2^2 \omega^2 + \dots$$

$$= \sum_x \frac{1}{2} m_x r_x^2 \omega^2$$

Taking the common factor  $\frac{1}{2}\omega^2$  out of the expression gives:

$$K = \frac{1}{2} \omega^2 (m_1 r_1^2 + m_2 r_2^2 + \dots)$$

$$= \frac{1}{2} \omega^2 (\sum_x m_x r_x^2)$$

The quantity in parenthesis, obtained by multiplying the mass of each particle by the square of the distance from the axis of rotation and adding these products, is called the Moment of Inertia of the body, denoted by  $I$ :

$$I = m_1 r_1^2 + m_2 r_2^2 + \dots = \sum_x m_x r_x^2$$

In terms of the moment of inertia ( $I$ ), the rotational kinetic energy ( $K$ ) of a rigid body is

$$K = \frac{1}{2} I \omega^2$$

- d. **Angular Momentum.** Angular Momentum ( $L$ ) is defined as the product of Moment of Inertia and Angular Velocity, ie  $L = I\omega$ .
- e. **Gyro Axes.** In gyro dynamics it is convenient to refer to the axis about which the torque is applied as the input axis and that axis about which the precession takes place as the output axis. The third axis, the spin axis, is self-evident. The  $XX^1$ ,  $YY^1$  and  $ZZ^1$  axes shown in the diagrams are not intended to represent the  $x, y$  and  $z$  axes of an aircraft in manoeuvre. However, if the  $XX^1$  (rotational) axis of the gyro is aligned with the direction of flight, the effects of flight manoeuvre on the gyro may be readily demonstrated in similar fashion to the instrument descriptions in Volume 5, Chapter 20.

## Classification of Gyroscopes

5. Conventional gyroscopes are classified in Table 1 in terms of the quantity they measure, namely:
  - a. **Rate Gyroscopes.** Rate gyroscopes measure the rate of angular displacement of a vehicle.
  - b. **Rate-integrating Gyroscopes.** Rate-integrating gyroscopes measure the integral of an input with respect to time.
  - c. **Displacement Gyroscopes.** Displacement gyroscopes measure the angular displacement from a known datum.

**Table 1 Classification of Gyros**

Type of Gyro	Uses in Guidance and Control	Gyro Characteristics
Rate Gyroscope	Aircraft Instruments	Modified single-degree-of-freedom gyro.
Rate-integrating Gyroscope	Older IN Systems	Modified single-degree-of-freedom gyro. Can also be a two-degree-of-freedom gyro.
Displacement Gyroscope	Heading Reference Older IN Systems Aircraft Instruments	Two degrees of freedom. Defines direction with respect to space, thus it is also called a space gyro, or free gyro.

6. It should be realized, however, that the above classification is one of a number of ways in which gyroscopes can be classified. Referring to Table 1, it will be seen that a displacement gyroscope could be classified as a two-degrees-of-freedom gyro or a space gyro. Note also that the classification of Table 1 does not consider the spin axis of a gyroscope as a degree of freedom. In this chapter, a degree of freedom is defined as the ability to measure rotation about a chosen axis.

## LAWS OF GYRODYNAMICS

### Rigidity in Space

7. If the rotor of a perfect displacement gyroscope is spinning at constant angular velocity, and therefore constant angular momentum, no matter how the frame is turned, no torque is transmitted to the spin axis. The law of conservation of angular momentum states that the angular momentum of a body is unchanged unless a torque is applied to that body. It follows from this that the angular momentum of the rotor must remain constant in magnitude and direction. This is simply another way of saying that the spin axis continues to point in the same direction in inertial space. This property of a gyro is defined in the First Law of Gyrodynamics.

### The First Law of Gyrodynamics

8. The first law of gyro dynamics states that:
 

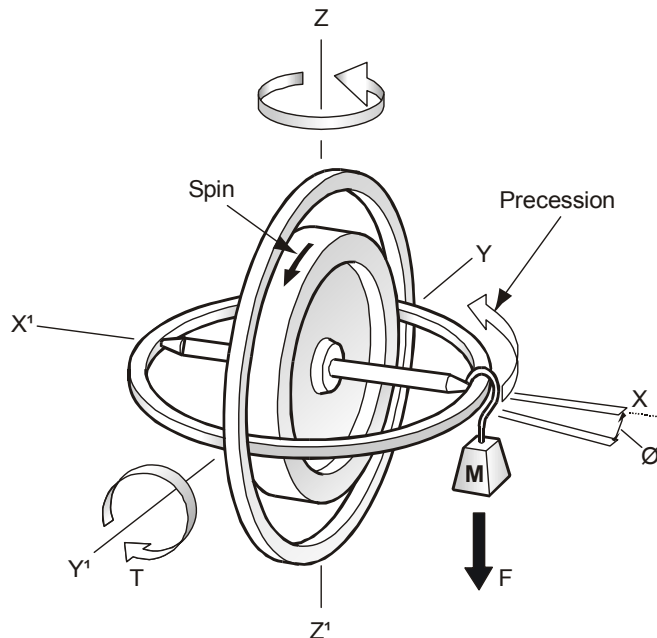
"If a rotating body is so mounted as to be completely free to move about any axis through the centre of mass, then its spin axis remains fixed in inertial space however much the frame may be displaced."
9. A space gyroscope loses its property of rigidity in space if the spin axis is subjected to random torques, some causes of which will be examined later.

## Precession

10. Consider the free gyroscope in Fig 3, spinning with constant angular momentum about the  $XX^1$  axis. If a small mass  $M$  is placed on the inner gimbal ring, it exerts a downward force  $F$  so producing a torque  $T$  about the  $YY^1$  axis. By the laws of rotating bodies, this torque should produce an angular acceleration about the  $YY^1$  axis, but this is not the case:

- Initially, the gyro spin axis will tilt through a small angle ( $\emptyset$  in Fig 3), after which no further movement takes place about the  $YY^1$  axis. The angle  $\emptyset$  is proportional to  $T$  and is a measure of the work done. Its value is almost negligible and will not be discussed further.
- The spin axis then commences to turn at a constant angular velocity about the axis perpendicular to both  $XX^1$  and  $YY^1$ , ie the  $ZZ^1$  axis. This motion about the  $ZZ^1$  axis is known as precession, and is the subject of the Second Law of Gyrodynamics.

13-33 Fig 3 Precession



## The Second Law of Gyrodynamics

11. The second law of gyrodynamics states that:

"If a constant torque ( $T$ ) is applied about an axis perpendicular to the spin axis of an unconstrained, symmetrical spinning body, then the spin axis will precess steadily about an axis mutually perpendicular to the spin axis and the torque axis. The angular velocity of precession ( $\Omega$ ) is given by  $\Omega = \frac{T}{I\omega}$ ."

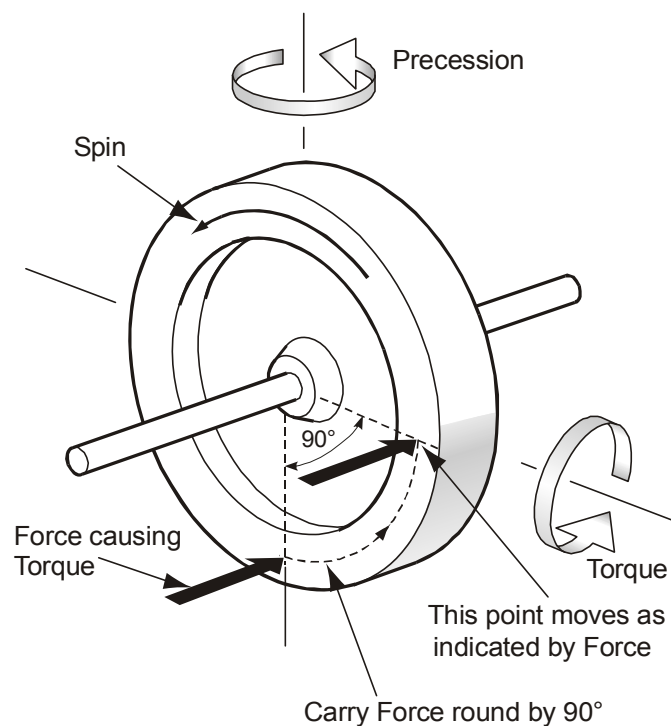
12. Precession ceases as soon as the torque is withdrawn, but if the torque application is continued, precession will continue until the direction of spin is the same as the direction of the applied torque. If, however, the direction of the torque applied about the inner gimbal axis moves as the rotor precesses, the direction of spin will never coincide with the direction of the applied torque.

## Direction of Precession

13. Fig 4 shows a simple rule of thumb to determine the direction of precession:

- a. Consider the torque as being due to a force acting at right angles to the plane of spin at a point on the rotor rim.
- b. Carry this force around the rim through  $90^\circ$  in the direction of rotor spin.
- c. The torque will apparently act through this point and the rotor will precess in the direction shown.

**13-33 Fig 4 Determining Precession**



## CONSERVATION OF ANGULAR MOMENTUM

### Explanation

14. In linear motion, if the mass is constant, changes in momentum caused by external forces will be indicated by changes in velocity. Similarly, in rotary motion, if the moment of inertia is constant, then the action of an external torque will be to change the angular velocity in speed or direction and, in this way, change the angular momentum. If, however, internal forces (as distinct from external torques) act to change the moment of inertia of a rotating system, then the angular momentum is unaffected. Angular momentum is the product of the moment of inertia and angular velocity, and if one is decreased so the other must increase to conserve angular momentum. This is the Principle of Conservation of Angular Momentum.

15. Consider the ice-skater starting her pirouette with arms extended. If she now retracts her arms she will be transferring mass closer to the axis of the pirouette, so reducing the radius of gyration. If the angular momentum is to be maintained then, because of the reduction of moment of inertia, the rate of her pirouette must increase, therefore:

- a. If the radius of gyration of a rotating body is increased, a force is considered to act in opposition to the rotation caused by the torque, decreasing the angular velocity.
- b. If the radius of gyration is decreased, a force is considered to act assisting the original rotation caused by the torque, so increasing the angular velocity.

### Cause of Precession

16. Consider the gyroscope rotor in Fig 5a spinning about the  $XX^1$  axis and free to move about the  $YY^1$  and  $ZZ^1$  axes. Let the quadrants (1, 2, 3 and 4) represent the position of the rotor in spin at one instant during the application of an external force to the spin axis, producing a torque about the  $YY^1$  axis. This torque is tending to produce a rotation about the  $YY^1$  axis while at the same instant the rotor spin is causing particles in quadrants 1 and 3 to recede from the  $YY^1$  axis, increasing their moment of inertia about this axis, and particles in quadrants 2 and 4 to approach the  $YY^1$  axis decreasing their moment of inertia about this axis. Particles in quadrants 1, 2, 3 and 4 tend to conserve angular momentum about  $YY^1$ , therefore:

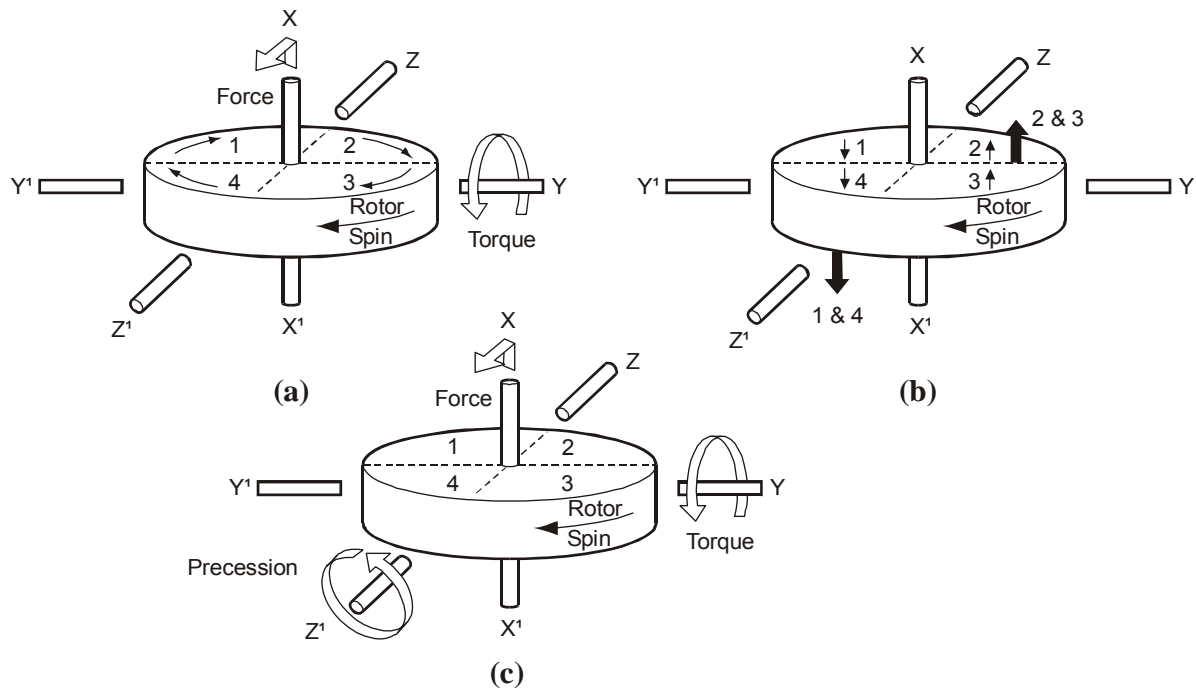
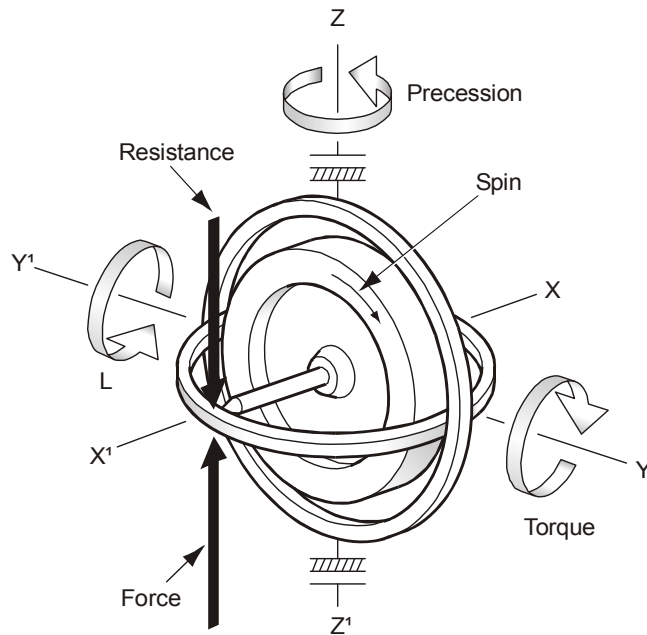
- a. Particles in quadrants 1 and 3 exert forces opposing their movement about  $YY^1$ .
- b. Particles in quadrants 2 and 4 exert forces assisting their movement about  $YY^1$ .

17. Hence, 1 and 4 exert forces on the rotor downwards, whilst 2 and 3 exert forces upwards. These forces can be seen to form a couple about  $ZZ^1$ , (Fig 5b), causing the rotor to precess in the direction shown in Fig 5c.

### Gyroscopic Resistance

18. In demonstrating precession, it was stated that, after a small deflection about the torque axis, movement about this axis ceased, despite the continued application of the external torque. This state of equilibrium means that the sum of all torques acting about this axis is zero. There must, therefore, be a resultant torque  $L$ , acting about this axis which is equal and opposite to the external torque, as shown in Fig 6. This resistance is known as Gyroscopic Resistance and is created by internal couples in a precessing gyroscope.

19. Consider now the gyroscope in Fig 5c spinning about an axis  $XX^1$  and precessing about the  $ZZ^1$  axis under the influence of a torque  $T$ , about the  $YY^1$  axis. The rotor quadrants represent an instant during the precession and spin. Using the argument of para 16, the particles in quadrants 1 and 3 are approaching the  $ZZ^1$  axis and exerting forces acting in the direction of precession, while in quadrants 2 and 4 the particles are receding from the  $ZZ^1$  axis and exerting forces in opposition to the precession. The resultant couple is therefore acting about the  $YY^1$  axis in opposition to the external torque. This couple is the Gyroscopic Resistance. It has a value equal to the external torque thus preventing movement about the  $YY^1$  axis.

**13-33 Fig 5 Instant of Spin and Precession****13-33 Fig 6 Gyroscopic Resistance**

20. Gyroscopic Resistance is always accompanied by precession, and it is of interest to note that, if precession is prevented, gyroscopic torque cannot form and it is as easy to move the spin axis when it is spinning as when it is at rest. This can be demonstrated by applying a torque to the inner gimbal of a gyroscope with one degree of freedom. With the  $ZZ^1$  axis locked, the slightest touch on the inner gimbal will set the gimbal ring (and the rotor) moving.

## Secondary Precession

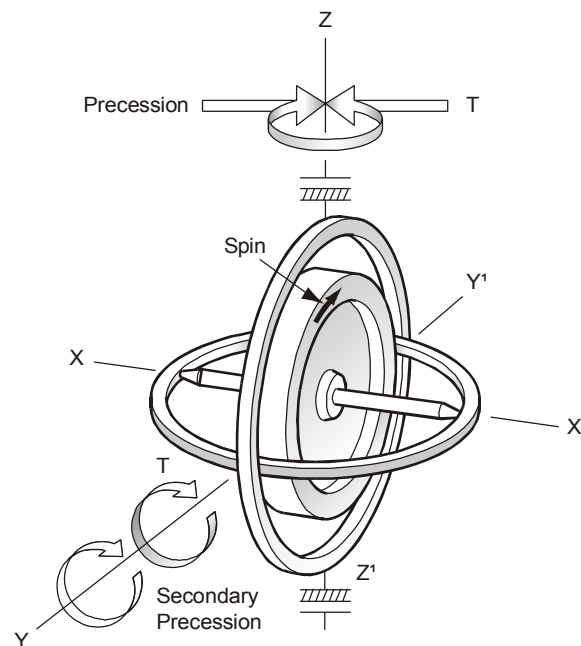
21. If a sudden torque is applied about one of the degrees of freedom of a perfect displacement gyroscope the following phenomena should be observed:

- Nodding or nutation occurs. Here it is sufficient to note that nutation occurs only for a limited period of time and eventually will cease completely. Additionally, nutation can only occur with a two-degree-of-freedom gyro and, to a large extent, it can be damped out by gyro manufacturers.
- A deflection takes place about the torque axis, (dip), which remains constant provided that the gyro is perfect and the applied torque is also constant.
- The gyro precesses, or rotates, about the  $ZZ^1$  axis.

22. If, however, an attempt is made to demonstrate this behaviour, it will be seen that the angle of dip will increase with time, apparently contradicting sub-para 21b.

23. To explain this discrepancy, consider Fig 7. If the gyro is precessing about the  $ZZ^1$  axis, some resistance to this precession must take place due to the friction of the outer gimbal bearings. If this torque  $T$  is resolved using the rule of thumb given in para 13, it will be seen that the torque  $T$  causes the spin axis to dip through a larger angle. This precession is known as secondary precession.

**13-33 Fig 7 Precession Opposed by Secondary Precession**

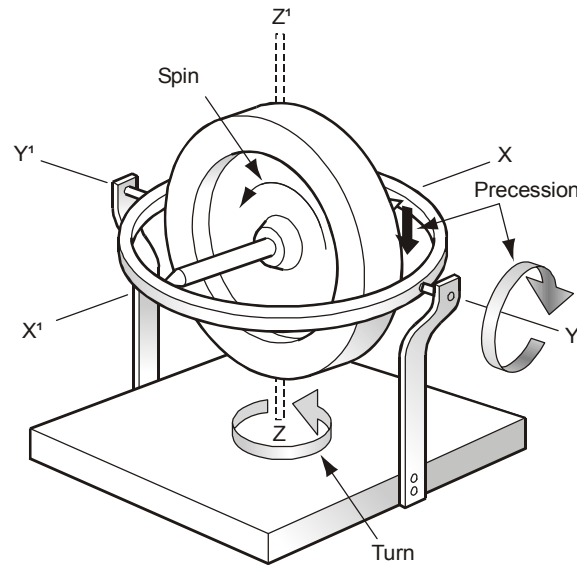


24. Secondary precession can only take place when the gyro is already precessing, thus its name. Note also that secondary precession acts in the same direction as the originally applied torque.

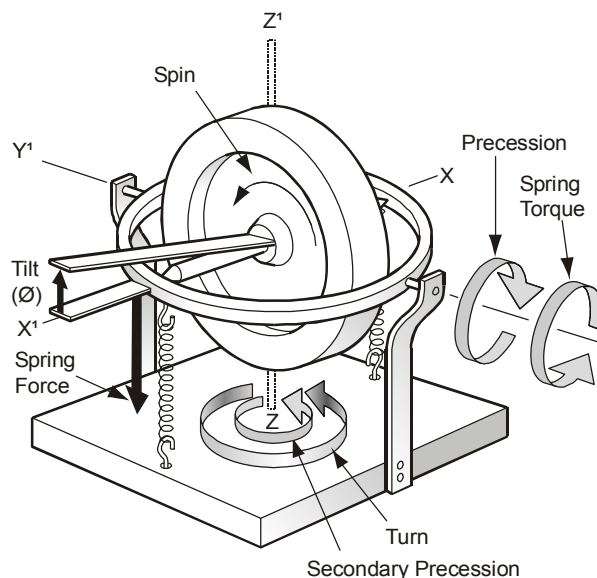
## THE RATE GYROSCOPE

### Principle of Operation

25. Fig 8 shows a gyroscope with freedom about one axis  $YY^1$ . If the frame of the gyro is turned about an axis  $ZZ^1$  at right angles to both  $YY^1$  and  $XX^1$ , then the spin axis will precess about the  $YY^1$  axis. The precession will continue until the direction of rotor spin is coincident with the direction of the turning about  $ZZ^1$ .

**13-33 Fig 8 Gyro with One degree of Freedom – Precession**

26. Suppose the freedom of this gyroscope about the gimbal axis is restrained by the springs connecting the gimbal ring to the frame as in Fig 9. If the gyroscope is now turned about the  $ZZ'$  axis, precession about the  $YY'$  axis is immediately opposed by a torque applied by the springs. It has been shown that any torque opposing precession produces a secondary precession in the same direction as the original torque (see para 24). If the turning of the frame is continued at a steady rate, the precession angle about the  $YY'$  axis will persist, distending one spring and compressing the other, thereby increasing the spring torque. Eventually, the spring torque will reach a value where it is producing secondary precession about  $ZZ'$  equal to, and in the same direction as, the original turning. When this state is reached, the gyroscope will be precessing at the same rate as it is being turned and no further torque will be applied by the turning. Any change in the rate of turning about the  $ZZ'$  axis will require a different spring torque to produce equilibrium, thus the deflection of the spin axis ( $\emptyset$  in Fig 9) is a measure of the rate of turning. Such an arrangement is known as a Rate Gyroscope, and its function is to measure a rate of turn, as in the Rate of Turn Indicator.

**13-33 Fig 9 Rate Gyroscope**



27. The relationship between the deflection angle and rate of turn is derived as follows:

Spring Torque is proportional to  $\phi$  or

Spring Torque =  $K\phi$  (where K is a constant)

At equilibrium:

Rate of Secondary Precession = Rate of Turn

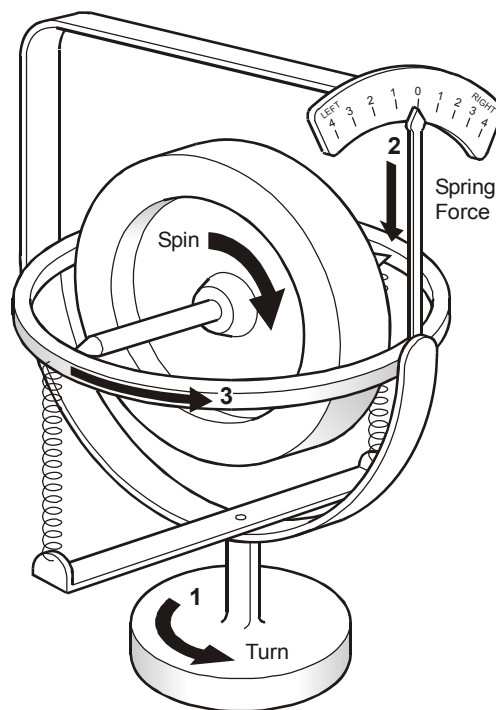
$$\text{ie } \frac{K\phi}{I\omega} = \text{Rate of Turn}$$

$\therefore \phi$  is proportional to Rate of Turn  $\times I\omega$

( $I\omega$  is the angular momentum of the rotor and is therefore constant).

The angle of deflection can be measured by an arrangement shown at Fig 10 and the scale calibrated accordingly.

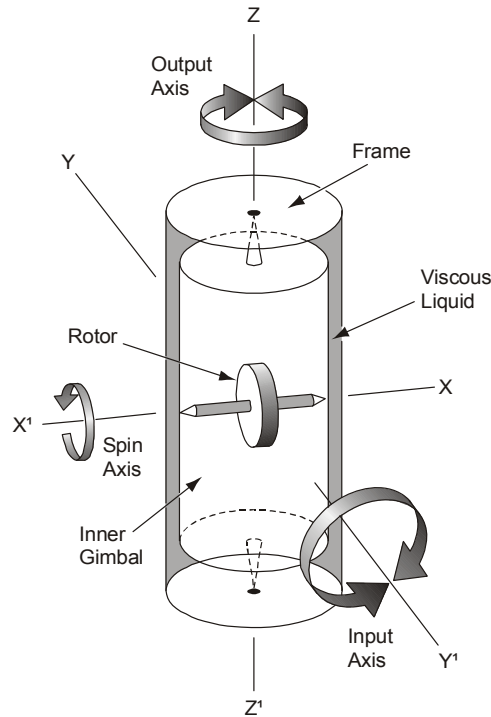
**13-33 Fig 10 Rate of Turn Indicator**



## THE RATE-INTEGRATING GYROSCOPE

### Principle of Operation

28. A rate-integrating gyroscope is a single degree of freedom gyro using viscous restraint to damp the precessional rotation about the output axis. The rate-integrating gyro is similar to the rate gyro except that the restraining springs are omitted and the only factor opposing gimbal rotation about the output axis is the viscosity of the fluid. Its main function is to detect turning about the input axis ( $YY^1$  in Fig 11), by precessing about its output axis ( $ZZ^1$  in Fig 11).

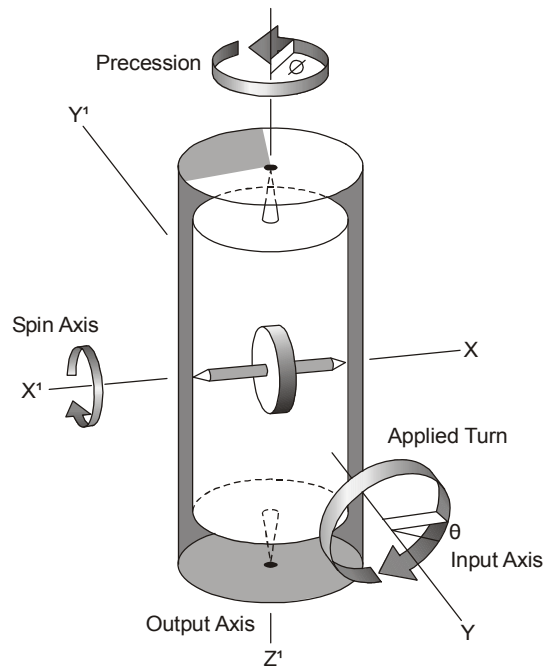
**13-33 Fig 11 Simple Rate-integrating Gyroscope**

29. The rate-integrating gyro was designed for use on inertial navigation stable platforms, where the requirement was for immediate and accurate detection of movement about three mutually perpendicular axes. Three rate-integrating gyros are used, each performing its functions about one of the required axes. These functions could be carried out by displacement gyros, but the rate-integrating gyro has certain advantages over the displacement type. These are:

- a. A small input rate causes a large gimbal deflection (gimbal gain).
- b. The gyro does not suffer from nutation.

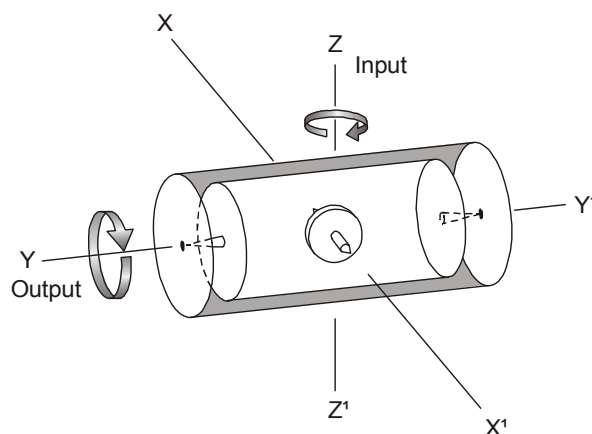
30. Fig 11 shows a simple rate-integrating gyro. It is basically a can within which another can (the inner gimbal) is pivoted about its vertical ( $ZZ^1$ ) axis. The outer can (frame) is filled with a viscous fluid which supports the weight of the inner gimbal so reducing bearing torques. The rotor is supported with its spin ( $XX^1$ ) axis across the inner gimbal. In a conventional non-floated gyro, ball bearings support the entire gimbal weight and define the output axis. In the floated rate-integrating gyro the entire weight of the rotor and inner gimbal assembly is supported by the viscous liquid, thereby minimizing frictional forces at the output ( $ZZ^1$ ) axis pivot points. The gimbal output must, however, be defined and this is done by means of a pivot and jewel arrangement. By utilizing this system for gimbal axis alignment, with fluid to provide support, the bearing friction is reduced to a very low figure.

31. The gyroscope action may now be considered. If the whole gyro in Fig 12 is turned at a steady rate about the input axis ( $YY^1$ ), a torque is applied to the spin axis causing precession about the output axis ( $ZZ^1$ ). The gimbal initially accelerates (precesses) to a turning rate such that the viscous restraint equals the applied torque. The gimbal then rotates at a steady rate about  $ZZ^1$ , proportional to the applied torque. The gyro output (an angle or voltage) is the summation of the amount of input turn derived from the rate and duration of turn and is therefore the integral of the rate input. (Note that the rate gyro discussed in paras 25 to 27 puts out a rate of turn only). The movement about the output axis may be made equal to, less than, or greater than movements about the input axis by varying the viscosity of the damping fluid. By design, the ratio between the output angle ( $\phi$ ) and the input angle ( $\theta$ ) can be arranged to be of the order of 10 to 1. This increase in sensitivity is called gimbal gain.

**13-33 Fig 12 Function of Rate-integrating Gyroscope**

32. A gyro mounted so that it senses rotations about a horizontal input axis is known as a levelling gyro. Two levelling gyros are required to define a level plane. Most inertial platforms using conventional gyros align the input axis of their levelling gyros with True North and East.

33. Motion around the third axis, the vertical axis, is measured by an azimuth gyro, ie one in which the input axis is aligned with the vertical, as in Fig 13.

**13-33 Fig 13 Rate-integrating Azimuth Gyroscope**

## THE DISPLACEMENT GYROSCOPE

### Definition

34. A displacement gyro is a two-degree-of-freedom gyro. It can be modified for a particular task, but it always provides a fixed artificial datum about which angular displacement is measured.

## Wander

35 Wander is defined as any movement of the spin axis away from the reference frame in which it is set.

36. **Causes of Wander.** Movement away from the required datum can be caused in two ways:

a. Imperfections in the gyro can cause the spin axis to move physically. These imperfections include such things as friction and unbalance. This type of wander is referred to as real wander since the spin axis is actually moving. Real wander is minimized by better engineering techniques.

b. A gyro defines direction with respect to inertial space, whilst the navigator requires Earth directions. In order to use a gyro to determine directions on Earth, it must be corrected for apparent wander due to the fact that the Earth rotates or that the gyro may be moving from one point on Earth to another (transport wander).

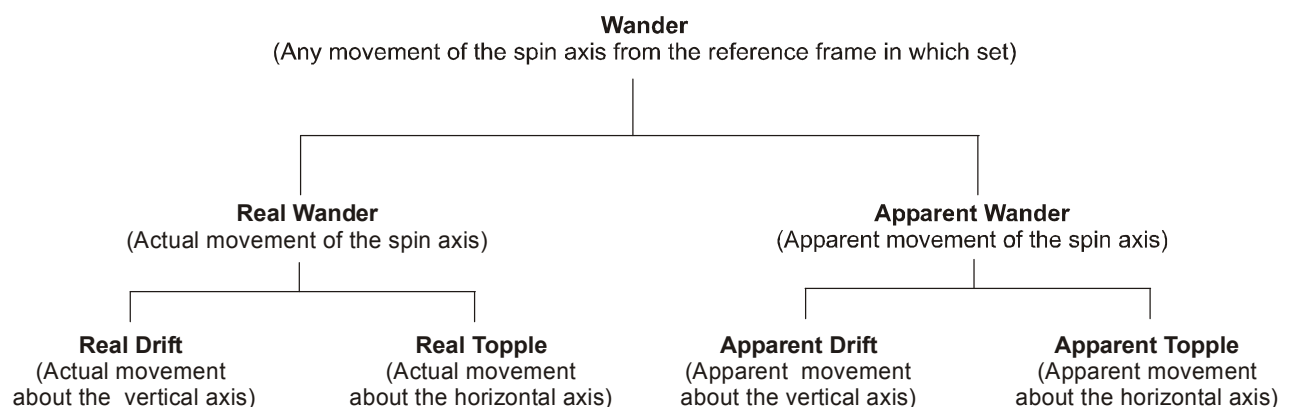
37. **Drift and Topple.** It is more convenient to study wander by resolving it into two components:

a. **Drift.** Drift is defined as any movement of the spin axis in the horizontal plane around the vertical axis.

b. **Topple.** Topple is defined as any movement of the spin axis in the vertical plane around a horizontal axis.

38. **Summary.** Table 2 summarizes the types of wander. From para 36 it should be apparent that the main concern when using a gyro must be to understand the effects of Earth rotation and transport wander on a gyro.

**Table 2 Types of Wander**



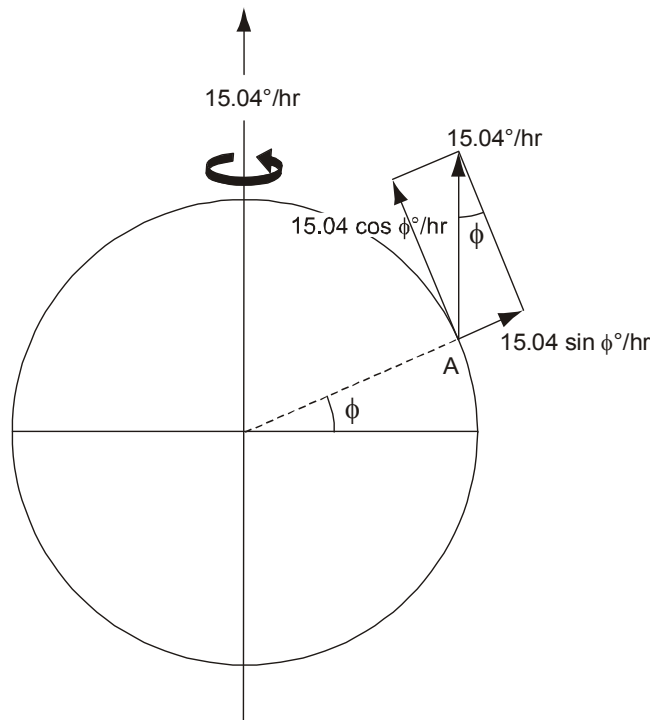
## Earth Rotation

39. In order to explain the effects of Earth rotation on a gyro it is easier to consider a single-degree-of-freedom gyro, since it has only one input and one output axis. The following explanation is based on a knowledge of rotational vector notation.

40. Consider a gyro positioned at a point A in Fig 14. It would be affected by Earth rotation according to how its input axis was aligned, namely:

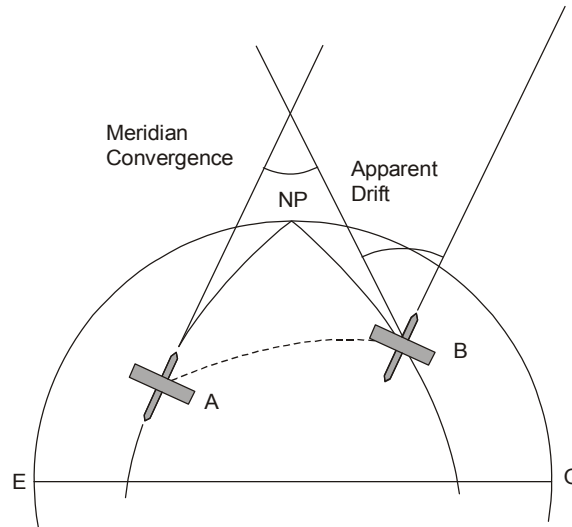
- If its input axis was aligned with the Earth's spin axis, it would detect Earth rate ( $\Omega_e$ ) of  $15.04^\circ/\text{hr}$ .
- Azimuth Gyro.** If its input axis was aligned with the local vertical it would detect  $15.04 \times \sin \phi^\circ/\text{hr}$ , where  $\phi$  = latitude. Note that, by definition, this is drift.
- North Sensitive Levelling Gyro.** If its input axis were aligned with local North, it would detect  $15.04 \times \cos \phi^\circ/\text{hr}$ . Note that, by definition, this is topple.
- East Sensitive Levelling Gyro.** Finally, if the input axis were aligned with local East, that is, at right angles to the Earth rotation vector, it would not detect any component of Earth rotation.

13-33 Fig 14 Components of Earth Rate



### Transport Wander

41. If an azimuth gyro spin axis is aligned with local North (i.e. the true meridian) at A in Fig 15 and the gyro is then transported to B, convergence of the meridians will make it appear that the gyro spin axis has drifted. This apparent drift is in addition to that caused by Earth rotation. The gyro has not in fact drifted; it is the direction of the True North which has changed. However, if the gyro is transported North-South, there is no change in the local meridian and therefore, no apparent drift. Similarly, as all meridians are parallel at the Equator, an East-West movement there produces no apparent drift. Transport rate drift thus depends on the convergence of the meridians and the rate of crossing them; i.e. the East-West component of ground speed ( $U$ ). The amount of convergence between two meridians ( $C$ ) is  $\text{ch long} \times \sin \text{lat}$ . Any given value of  $U$  thus produces an increase in apparent gyro drift as latitude increases.

**13-33 Fig 15 Apparent Drift**

The amount of drift due to transport rate may be found as follows:

$$C (^{\circ}/\text{hr}) = [\text{ch long/hr}] \times \sin \phi.$$

$$\text{Now, ch long/hr} = \frac{\text{ch Eastings (nm / hr)}}{60} \times \sec \phi$$

and, since  $1^{\circ} = 60 \text{ nm}$  and  $\text{ch Eastings (nm/hr)} = U$

$$C = \frac{U}{60} \times \sec \phi \times \sin \phi (^{\circ}/\text{hr})$$

$$\text{but, } \sec \phi \times \sin \phi = \frac{1}{\cos \phi} \times \sin \phi$$

$$= \frac{\sin \phi}{\cos \phi} = \tan \phi$$

$$\therefore C = \frac{U}{60} \times \tan \phi (^{\circ}/\text{hr})$$

This can be converted to radians/hour by multiplying by  $\frac{\pi}{180}$

$$\therefore C = \frac{U}{60} \times \tan \phi \times \frac{\pi}{180} = U \times \tan \phi \times \frac{\pi}{60 \times 180}$$

Now an arc of length 60 nm on the Earth's surface subtends an angle of  $1^{\circ}$  ( $\pi/180^{\circ}$ ) at the centre of the Earth

$$\therefore R \times \frac{\pi}{180} = 60 \text{ where } R = \text{Earth's radius}$$

$$\text{or, } \frac{1}{R} = \frac{\pi}{60 \times 180}$$

Substituting into the above equation for Meridian Convergence (radians/hour)

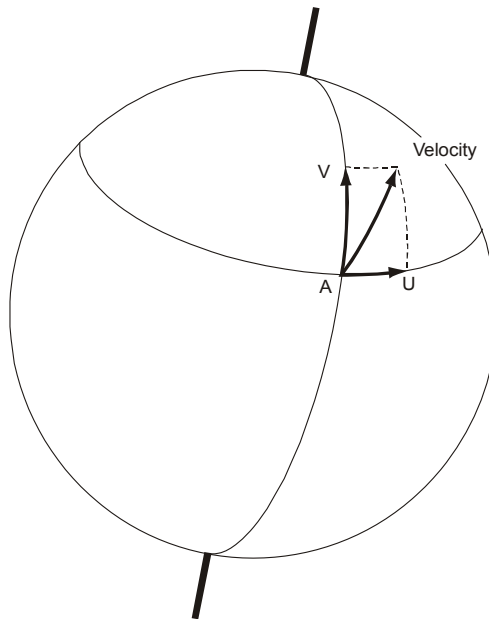
$$C = U \times \tan \phi \times \frac{1}{R}$$

$$\text{or, } C = \frac{U}{R} \times \tan \phi \text{ (radians/hour)}$$

42. Consider now two levelling gyros, whose input axes are North and East respectively, and whose output axes are vertical.

- The East component of aircraft velocity in Fig 16 will be sensed by the North gyro as a torque of  $\frac{U}{R}$  about its input axis. If the gyro is not corrected for this transport wander, it is said, by definition, to topple.
- Similarly, due to the effect of aircraft velocity North, the East gyro will topple at the rate of  $\frac{V}{R}$ .

**13-33 Fig 16 Transport Wander**



#### Apparent Wander Table

43. All of the equations derived in the study of Earth rate and transport wander rate are summarized in Table 3. The units for Earth rate can be degrees or radians, whilst for transport wander they are radians.

44. **Correction Signs.** The correction signs of Table 3 apply only to the drift equations, and they should be applied to the gyro readings to obtain true directions. These correction signs will be reversed for the Southern Hemisphere.

**Table 3 Components of Drift and Topple – Earth Rate and Transport Wander Rate**

	Input Axis Alignment			Correction Sign
	Local North	Local East	Local Vertical	
<b>Earth Rate</b> degrees (or radians) per hour	$\Omega_e \cos \phi$	Nil	$\Omega_e \sin \phi$	+
<b>Transport Wander</b> radians per hour	$\frac{U}{R}$	$\frac{-V}{R}$	$\frac{U}{R} \tan \phi$	+E -W
	Topple		Drift	

$\Omega_e$  = Angular Velocity of the Earth

$R$  = Earth's Radius

$\phi$  = Latitude

$U$  = East/West component of groundspeed

$V$  = North/South component of groundspeed

### Practical Corrections for Topples and Drift

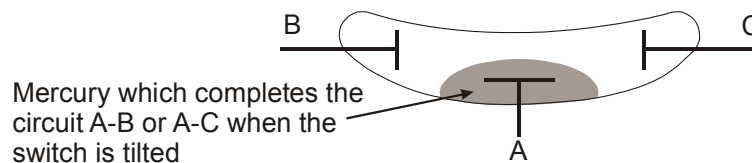
45. If all the corrections of Table 3 were applied to three gyros with their input axes aligned to true North, true East and the local vertical, true directions would be defined continuously, and in effect the gyros would have been corrected for all apparent wander. However, these corrections make no allowance for the real wander of a gyro and consequently an error growth proportional to the magnitude of the real drift and topple will exist. As a rough rule of thumb, an inertial platform employing gyros with real drift rates in the order of  $0.01^\circ/\text{hr}$  will have a system error growth of 1 to 2 nm/hr CEP.

46. Flight instruments, on the other hand, employ cheaper, lower quality gyros whose drift rates may be in the order of  $0.1^\circ/\text{hr}$ . If these real drift rates were not compensated for, system inaccuracies would be unacceptably large. For this reason, some flight instruments make use of the local gravity vector to define the level plane, thus compensating for both real and apparent drifts.

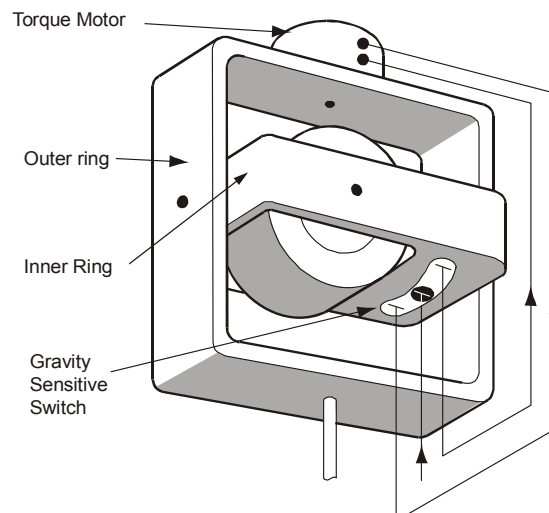
47. Specifically, gyro wander may be corrected in the following ways:

- a. **Topple.** Topple is normally corrected for in gyros by the use of either gravity switches (see Figs 17 and 18), or by case levelling devices (see Fig 19). These devices sense movement away from the vertical, and send appropriate signals to a torque motor until the vertical is re-established. The levelling accuracy of these methods is approximately  $1^\circ$ .
- b. **Drift.** Drift corrections can be achieved by:
  - (1) Calculating corrections using Table 3 and applying them to the gyro reading.
  - (2) Applying a fixed torque to the gyro so that it precesses at a rate equal to the Earth rate for a selected latitude. Although this method is relatively simple, it has the disadvantage that the compensation produced will only be correct at the selected latitude.
  - (3) Applying variable torques, using the same approach as in (2) above, but being able to vary the torque according to the latitude. These azimuth drift corrections make no allowance for real drift, which can only be limited by coupling the azimuth gyro to a flux valve.

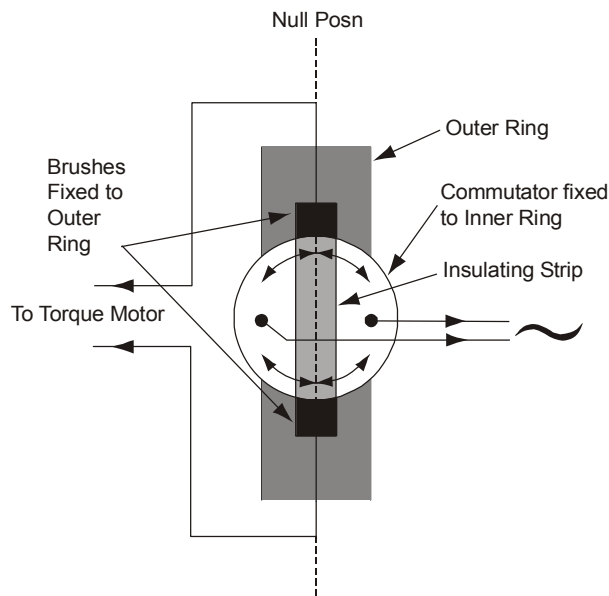
**13-33 Fig 17 Gravity Sensitive Switch**



**13-33 Fig 18 Gravity Levelling**





**13-33 Fig 19 Synchro Case Levelling Device**

48. To complete this study of the displacement gyro, it remains to mention a limitation and an error peculiar to this type of gyro, namely gimbal lock and gimbal error.

### **Gimbal Lock**

49. Gimbal lock occurs when the gimbal orientation is such that the spin axis becomes coincident with an axis of freedom. Effectively the gyro has lost one of its degrees of freedom, and any attempted movement about the lost axis will result in real wander. This is often referred to as toppling, although drift is also present.

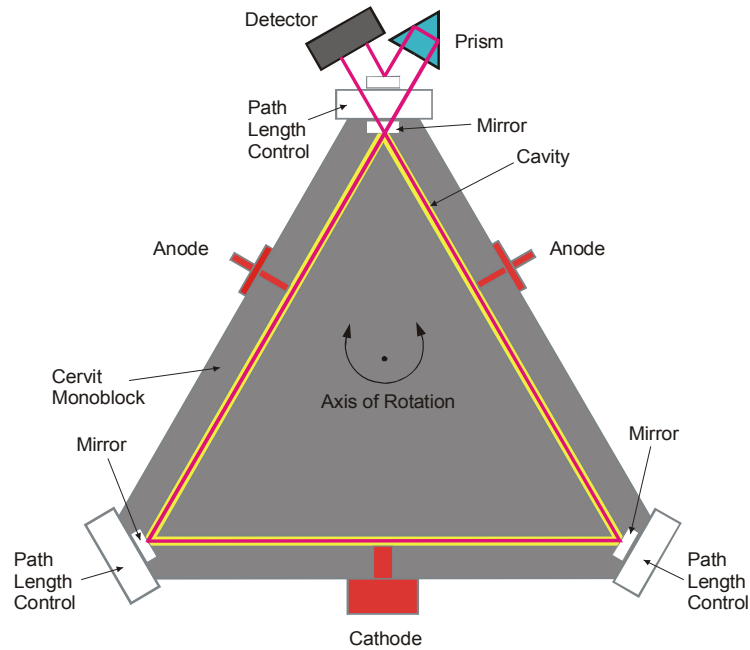
### **Gimbal Error**

50. When a 2-degree-of-freedom gyroscope with a horizontal spin axis is both banked and rolled, the outer gimbal must rotate to maintain orientation of the rotor axis, thereby inducing a heading error at the outer gimbal pick-off. The incidence of this error depends upon the angle of bank and the angular difference between the spin axis and the longitudinal axis and, as in most systems, the spin axis direction is arbitrary relative to North, the error is not easily predicted. Although the error disappears when the aircraft is levelled, it will have accumulated in any GPI equipment, producing a small error in computed position.

## **OTHER TYPES OF GYRO**

### **The Ring Laser Gyro**

51. **Introduction.** The ring laser gyro is one of the modern alternatives to conventional gyros for a number of applications including aircraft inertial navigation systems and attitude/heading reference systems. The ring laser gyro has no moving parts and is not a gyroscope in the normal meaning of the word. The ring laser gyro is, however, a very accurate device for measuring rotation and became the system of choice for use with strapdown inertial navigation systems. A schematic diagram of a ring laser gyro is at Fig 20.

**13-33 Fig 20 Schematic Diagram of a Ring Laser Gyro**

**52. Principle of Operation.** In ring laser gyros, the rotating mass of the conventional gyro is replaced by two contra-rotating beams of light. The main body of the gyro consists of a single piece (or 'monoblock') of a vitreous ceramic of low temperature coefficient (typically 'Cervit' or 'Zerodour'). A gas-tight cavity is accurately machined into the monoblock and this cavity is then filled with an inert gas, typically a mixture of Helium and Neon. A DC electrical discharge ionizes the gas and causes the lasing action. Two beams of light are produced, flowing in opposite directions in the cavity. Mirrors are used to reflect the beams around the enclosed area, producing a 'laser-in-a-ring' configuration. The frequency of oscillation of each beam corresponds to the cavity resonance condition. This condition requires that the optical path length of the cavity be an integral number of wavelengths. The frequency of each beam is therefore dependant on the optical path length.

**53. Effect of Movement.** At rest, the optical path length for each beam is identical; therefore, the frequencies of the two laser beams are the same. However, when the sensor is rotated about the axis perpendicular to the lasing plane, one beam travels an increased path length, whilst the other travels a reduced path length. The two resonant frequencies change to adjust to the longer or shorter optical path, and the frequency difference is directly proportional to the rotation rate. This phenomenon is known as the Sagnac effect. The frequency difference is measured by the beaming of an output signal for each wave on to photo detectors spaced one quarter of a wavelength apart, causing an optical effect known as an interference fringe. The fringe pattern moves at a rate that is directly proportional to the frequency difference between the two beams. It is converted to a digital output, where the output pulse rate is proportional to the input turn rate, and the cumulative pulse count is proportional to the angular change. This effect can be quantified using simplified maths, where it can be shown that the frequency difference  $\Delta f$  of the two waves is:

$$\Delta f = \frac{4A\Omega}{\lambda L}$$

Where

A is the area enclosed by the path.

$\lambda$  is the oscillating wavelength.

L is the length of the closed path.

$\Omega$  is the rate of rotation.

54. **Gyro Control.** The reason why the two beams have to occupy the same physical cavity is the sensitivity of laser light to cavity length. If they did not, a temperature induced difference in path length could result in a large frequency mismatch between the two beams. The path length control mechanism is used to alter the intensity of the laser and thus control expansion due to excess heat. To help avoid perceived differences in path length due to flow of the Helium/Neon gas mix, two anodes are used to balance any flow caused by ionization.

55. **Error Sources.** Ring laser gyros are subject to a number of errors, the most notable of which are:

a. **Null Shift.** Null shift arises due to a difference in path length as perceived by the two opposite beams, thus producing an output when no rotation exists. The major causes of perceived path length difference are:

(1) Differential movement of the gas in the cavity.

(2) Small changes in the refractive index of the monoblock material as the direction of travel of the laser light changes.

b. **Lock-in.** Lock-in occurs when the input rotation rate of the gyro is reduced below a critical value causing the frequency difference between the clockwise and anti-clockwise beams to drop to zero. One of the main causes of this phenomenon is backscatter of light at the mirrors. Some of the clockwise beam is reflected backwards, thereby contaminating the anti-clockwise beam with the clockwise frequency. Similarly, backscattering of the anti-clockwise beam contaminates the clockwise beam. With low input rotational rates, the two beams soon reach a common frequency which renders detection of rotation impossible. Several methods are used to ensure that lock-in is minimized. One method is to physically dither the gyro by inputting a known rotation rate in one direction, immediately followed by a rotation rate in the opposite direction. As the dither rate is known, it can be removed at the output stage. The dither ensures that the two frequencies are kept far enough apart to avoid lock-in.

56. **Advantages of the Ring Laser Gyro.** The main advantages of the ring laser gyro are:

- a. Its performance is unaffected by high 'g'.
- b. It has no moving parts and therefore has high reliability and low maintenance requirements.
- c. It has a rapid turn-on time.

57. **Disadvantages of the Ring Laser Gyro.** The technical problems associated with ring laser gyros can all be overcome. However, solution of these problems inevitably increases costs which are already very high due to the complex, 'clean-room' manufacturing facilities needed to provide:

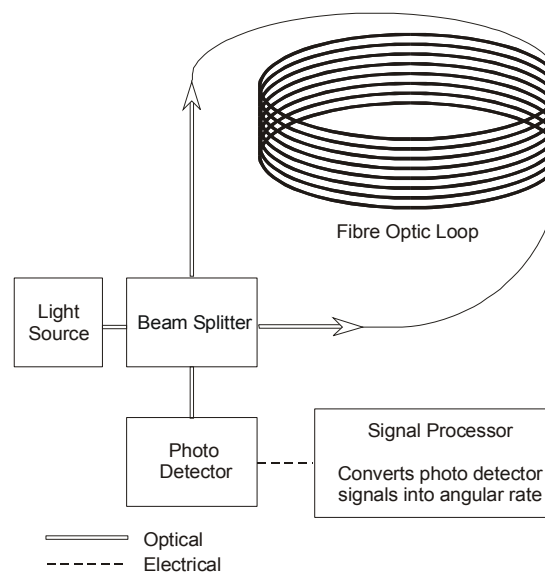
- a. Precision machining and polishing.
- b. High quality mirrors.
- c. Very good optical seals.
- d. A carefully balanced mix of Helium and Neon, free of contaminants.

58. **Summary.** While the ring laser gyro represents a major advance over the traditional spinning gyro, it is only one of a number of possible alternatives. The search for new gyroscopic devices continues, driven by considerations of both cost and accuracy.

### Fibre Optic Gyros

59. As previously outlined, the major disadvantage of ring laser gyros is their high cost due to the precise engineering facilities required to manufacture them. The fibre optic gyro (see Fig 21), first tested in 1975, works on the same principal as the ring laser gyro (the Sagnac effect) but no longer relies on a complex and costly block and mirror system since it uses a coil of fibre optic cable.

13-33 Fig 21 Fibre Optic Gyro

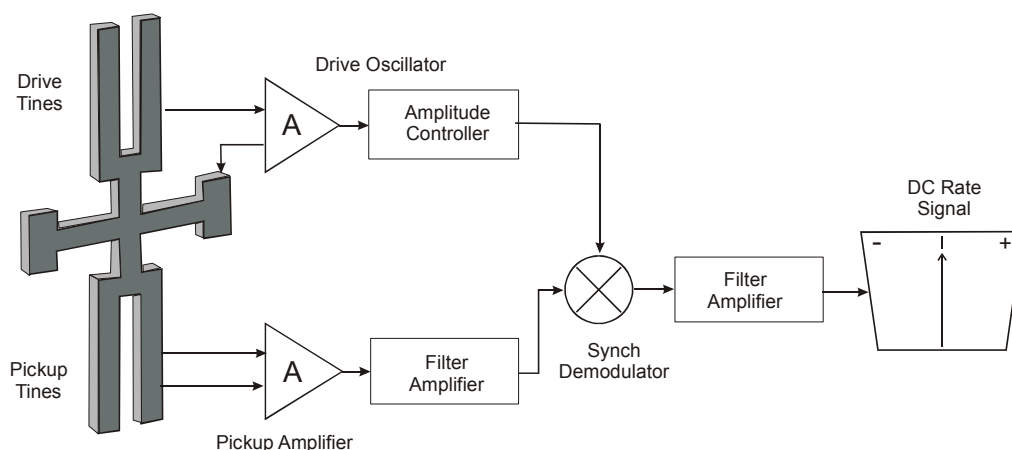


### Vibrating Gyros

60. Vibrating gyros work by exploiting the Coriolis effect and, while not yet as accurate as optical gyros, are much smaller and cheaper to produce.

61. The 'GyroChip' (Fig 22) uses a vibrating quartz tuning fork as a Coriolis sensor, coupled to a similar fork as a pickup to produce the rate output signal. The piezoelectric drive tines are driven by an oscillator to vibrate at a precise amplitude, causing the tines to move toward and away from one another at a high frequency. This vibration causes the drive fork to become sensitive to angular rate about an axis parallel to its tines, defining the true input axis of the sensor.

13-33 Fig 22 The 'GyroChip' Vibrating Gyro



62. Vibration of the drive tines causes them to act like the arms of a spinning ice skater, where moving them in causes the skater's spin rate to increase, and moving them out causes a decrease in rate. An applied rotation rate causes a sine wave of torque to be produced, resulting from 'Coriolis Acceleration', in turn causing the tines of the Pickup Fork to move up and down (not toward and away from one another) out of the plane of the fork assembly.

63. The pickup tines thus respond to the oscillating torque by moving in and out of plane, causing electrical output signals to be produced by the Pickup Amplifier. These signals are amplified and converted into a DC signal proportional to rate by use of a synchronous switch (demodulator) which responds only to the desired rate signals. The DC output signal of the 'GyroChip' is directly proportional to input rate, reversing sign as the input rate reverses, since the oscillating torque produced by Coriolis reverses phase when the input rate reverses.



# AP3456 The Central Flying School (CFS) Manual of Flying

Version 10.0 – 2018

Volume 14 – Electronics

AP3456 is sponsored by the Commandant Central Flying School

## Contents

Commandant CFS - Foreword  
Introduction and Copyright Information  
AP3456 Contact Details

### Chapter

14-1	Basic Electricity
14-2	Magnetism
14-3	Electromagnetic Induction and Inductance
14-4	Electrostatics and Capacitance
14-5	Measuring Instruments
14-6	DC Machines
14-7	Alternating Current Theory
14-8	AC Circuits
14-9	Selectivity of Tuned Circuits
14-10	AC Generators and Motors
14-11	Transformers
14-12	Fundamental Electronic Components
14-13	Power Supplies
14-14	Amplifiers
14-15	Oscillators
14-16	Transmitters
14-17	Receivers
14-18	Transmission Lines
14-19	Aerials
14-20	Microwaves
14-21	Radio Propagation
14-22	Telecommunications Principles
14-23	Introduction to Electronic Warfare (EW)
14-24	Electronic Support Measures (ESM)
14-25	Electronic Attack (EA)
14-26	Electronic Protection (EP)

## CHAPTER 1 - BASIC ELECTRICITY

Contents	Page
Electron Theory .....	2
Conductors and Insulators .....	2
Electric Units.....	2
Ohm's Law and Resistance.....	3
EMF and Potential Difference .....	3
Resistivity .....	4
Temperature Coefficient of Resistance .....	4
Resistors in Series.....	4
Resistors in Parallel.....	5
Kirchhoff's Laws .....	7
Rheostats and Potentiometers .....	7
Energy and Power .....	8
Heating Effect of a Current.....	8
The Maximum Power Transfer Theorem .....	8
Practical Resistors.....	9
Preferred Values of Resistance.....	10
The Resistor Colour Code.....	10

### Table of Figures

14-1 Fig 1 An Atom.....	2
14-1 Fig 2 Circuit to show EMF and PD .....	3
14-1 Fig 3 Resistors in Series .....	5
14-1 Fig 4 Resistors in Parallel .....	6
14-1 Fig 5 Kirchhoff's First Law .....	7
14-1 Fig 6 Kirchhoff's Second Law.....	7
14-1 Fig 7 A Rheostat.....	7
14-1 Fig 8 A Potentiometer .....	8
14-1 Fig 9 Construction of a Wire-wound Resistor .....	9
14-1 Fig 10 Two Types of Carbon Resistor .....	9
14-1 Fig 11 Examples of Resistors with their Values.....	11

### Table

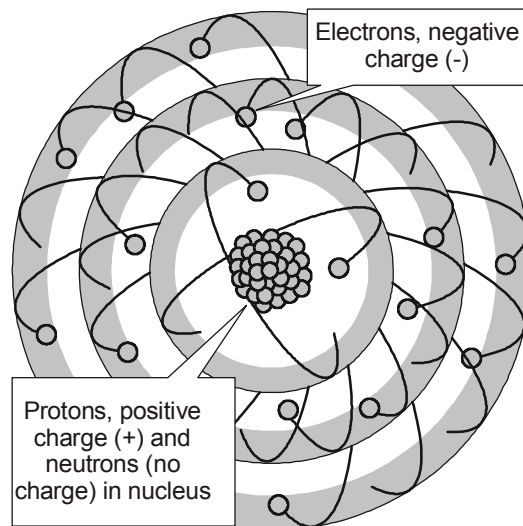
Table 1 Colour Code for Fixed Resistors .....	10
-----------------------------------------------	----



## Electron Theory

1. The smallest portion of any piece of matter which still retains the properties of the original substance is called a molecule. Molecules are made up of atoms which in turn are composed of a mixture of negatively and positively charged particles (see Fig 1). Normally there is an equal amount of each charge in an atom; the negatively charged particles (electrons) being balanced by the positively charge particles (protons) in the nucleus. The neutrons which form part of the nucleus have no charge.

14-1 Fig 1 An Atom



## Conductors and Insulators

2. In some materials, the outer orbital electrons are loosely attached to the atom's nucleus and are free to wander within the body of the material. If a voltage is applied across the material, these free electrons experience a force and will move under its influence. A bodily movement of electrons is called an electric current and substances in which appreciable amounts will flow are called conductors. Insulators will pass only a very small number of electrons, even at high voltages.

3. The property of a substance which tends to oppose the flow of electric current is called the resistance of the substance. Good conductors have low resistance while poor conductors (insulators) have high resistance.

## Electric Units

4. The fundamental unit of electricity is the electron, but this is far too small to be of any practical use for measurement. Consequently the practical unit of electricity or electric charge, the coulomb, is much larger. The coulomb is equal to the charge on  $6.28 \times 10^{18}$  electrons. When a charge of one coulomb flows in one second through a conductor, it constitutes a current of one ampere (usually shortened to 'amp').

$$\text{ie, amperes (A)} = \frac{\text{coulombs (Q)}}{\text{seconds (t)}}$$

5. Current will only flow in a conductor if the conducting material is connected between the terminals of a power source such as a battery. The battery is said to provide an electromotive force or emf. The practical unit of emf is the volt. The electromotive force may be thought of as driving electric current in the same way as pressure will drive a current of water through a pipe. Water flows from a greater height to a lower point and similarly electric charge is said to flow from a higher to a lower potential. The amount of water flowing per second will depend upon the diameter of the pipe. Likewise, the electric current will depend upon the resistance of the conducting wire, which in turn will depend on its diameter.

6. The direction of conventional current flow is taken as being from positive to negative, which is the opposite direction to that in which the electrons travel.

### Ohm's Law and Resistance

7. Ohm's law states that under constant physical conditions the current flowing in a conductor is directly proportional to the potential difference between its ends and inversely proportional to its resistance. If the potential difference is increased then the current will increase by the same factor; if resistance increases then the current decreases. Ohm's law is expressed by the following formula:

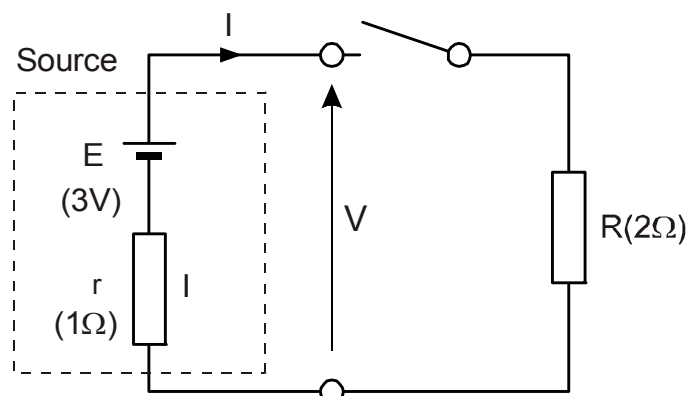
$$\text{Current (I)} = \frac{\text{Voltage (V)}}{\text{Resistance (R)}}$$

8. The unit of resistance is the Ohm, and a conductor is said to have a resistance of one ohm if it will allow a current of one ampere to flow when a potential difference of one volt is applied across it. The symbol for ohms is the Greek capital letter omega ( $\Omega$ ). For larger values of resistance, the units Kilohm and Megohm are used (Kilohm ( $K \Omega$ ) = 1,000  $\Omega$ ; Megohm ( $M \Omega$ ) = 1,000,000  $\Omega$ )

### EMF and Potential Difference

9. It is important to distinguish between electromotive force (emf) and potential difference (pd), because, although they are both measured in volts, they differ considerably. For any given power source the off-load voltage across the output terminals (pd) is equal to the driving force (emf) of the source. However, once an external load is connected across the terminals pd is no longer equal to emf. This is because all sources of power have internal resistance and voltage is dropped across this resistance once current flows in the circuit. To help distinguish between the two terms, emf is represented by the letter 'E' and pd by the letter 'V'. The following example is intended to illustrate the difference.

**14-1 Fig 2 Circuit to show EMF and PD**



Example 1. In Fig 2, the emf of the battery (E) is 3V and an internal resistance (r) of  $1\Omega$  is assumed. When the switch is closed, current (I) will flow through the external load (R) and the internal resistance of the battery,

$$\text{ie } I = \frac{E}{R + r} = \frac{3}{2 + 1} = 1\text{A}$$

and the voltage at the battery terminal

$$(V) = IR = 1 \times 2 = 2\text{V}.$$

Thus, it can be seen that although the battery emf is 3V, there is only a pd of 2V across the terminals.

### Resistivity

10. The longer a piece of wire is then the higher its resistance will be. Similarly, the thicker the wire is the lower its resistance will be. However, resistance is also determined by the nature of the material being used. This is referred to as the resistivity of the material. Resistivity is expressed in ohms metres and is defined as being the resistance per metre of material with cross-sectional area of one square metre; it is represented by the Greek letter rho ( $\rho$ ). Therefore:

$$\text{Resistance (R)} = \frac{\text{Resistivity } (\rho) \times \text{Length (l)}}{\text{Cross-sectional area (a)}}$$

### Temperature Coefficient of Resistance

11. In the statement of Ohm's law the words "under constant physical conditions" are used; this is because the resistance of most materials varies with temperature. For metallic conductors, resistance increases nearly uniformly with temperature, and each material is given a temperature coefficient. This coefficient is defined as the fractional increase in resistance per degree increase in temperature above  $0^\circ\text{C}$ . Temperature coefficient of resistance is represented by the Greek letter alpha ( $\alpha$ ). For practical purposes the following formula can be applied:

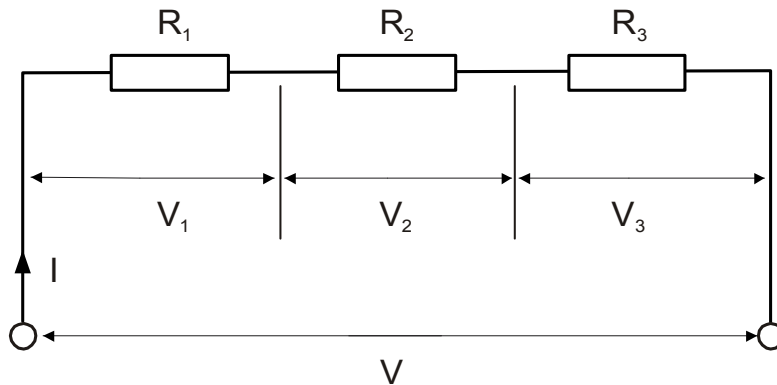
$$R_t = R_0 (1 + \alpha t)$$

Where  $R_t$  = resistance at  $t^\circ\text{C}$   
 $R_0$  = resistance at  $0^\circ\text{C}$   
 $\alpha$  = temperature coefficient of resistance  
 $t$  = operating temperature

12. The temperature coefficients of alloys are generally much smaller than those of pure metals, and special alloys are used in making standard resistors which vary only slightly with changes in temperature. The resistance of carbon and glass decreases with increases in temperature so these materials are given a negative temperature coefficient.

### Resistors in Series

13. When resistors are connected in series, their resultant resistance is equal to the sum of their separate values. In Fig 3, three resistors  $R_1$ ,  $R_2$ , and  $R_3$  are connected in series.

**14-1 Fig 3 Resistors in Series**

The voltages across the resistors are  $V_1$ ,  $V_2$ , and  $V_3$  respectively. Since the same current ( $I$ ) flows through each resistor, then by applying Ohm's law:

$$V_1 = IR_1 \quad V_2 = IR_2 \quad V_3 = IR_3$$

The total voltage across the three resistors is given by,

$$\begin{aligned} V_T &= V_1 + V_2 + V_3 \quad \text{or} \quad IR_1 + IR_2 + IR_3 \\ &= I(R_1 + R_2 + R_3) \end{aligned}$$

Therefore, the three resistors in series are equivalent to a single resistor  $R_T$  of value,

$$R_T = R_1 + R_2 + R_3$$

Example 2. Three resistors of values  $100 \, \Omega$ ,  $220 \, \Omega$  and  $470 \, \Omega$ , are connected in series. Calculate the total resistance, and find the current flow if  $50V$  is applied across the combination.

$$\begin{aligned} R_T &= R_1 + R_2 + R_3 \\ &= 100 \, \Omega + 220 \, \Omega + 470 \, \Omega \\ &= 790 \, \Omega \end{aligned}$$

Total resistance is  $790 \, \Omega$

$$\begin{aligned} I &= \frac{V}{R_T} \\ &= \frac{50V}{790\Omega} = 0.063 \, A \quad \text{or} \quad 63 \, mA \end{aligned}$$

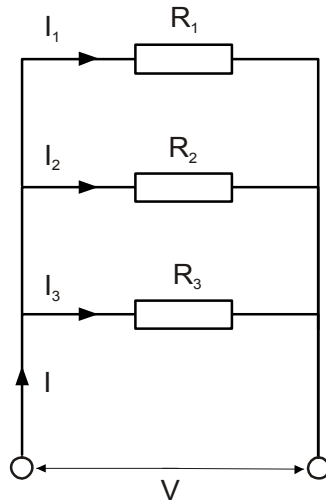
A current of  $63 \, mA$  flows through the circuit.

### Resistors in Parallel

14. When resistors are connected in parallel, their total resistance is given by the equation:

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

15. In Fig 4 the three resistors are connected in parallel. The same voltage appears across each resistor but the currents flowing through the resistors are not the same.

**14-1 Fig 4 Resistors in Parallel**

Referring to Fig 4, the currents flowing through the resistors are  $I_1$ ,  $I_2$  and  $I_3$ , respectively. Once again, by applying Ohm's law:

$$I_1 = \frac{V}{R_1} \quad I_2 = \frac{V}{R_2} \quad I_3 = \frac{V}{R_3}$$

The total current flowing in the circuit is given by,

$$I = I_1 + I_2 + I_3 = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3}$$

$$= V \left( \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right)$$

But,  $I = \frac{V}{R}$  where  $R$  is the equivalent of the three resistors. Dividing throughout by  $V$  gives:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

**Example 3.** Three resistors, of values  $100 \, \Omega$ ,  $200 \, \Omega$ , and  $400 \, \Omega$ , are connected in parallel. Calculate the total resistance and find the current flow if  $200V$  is applied across the combination.

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

$$= \frac{1}{100\Omega} + \frac{1}{200\Omega} + \frac{1}{400\Omega}$$

$$= \frac{4+2+1}{400\Omega} = \frac{7}{400\Omega}$$

$$R_T = \frac{400\Omega}{7} = 57.14 \, \Omega$$

The total resistance is  $57.14 \, \Omega$

$$I = \frac{V}{R_T} = \frac{200V}{57.14\Omega} = 3.5 \, A$$

The total current flowing in the circuit is  $3.5 \, A$ .

## Kirchhoff's Laws

16. In the preceding paragraphs on resistors in series and parallel, two important laws due to Kirchhoff have been assumed. These are:

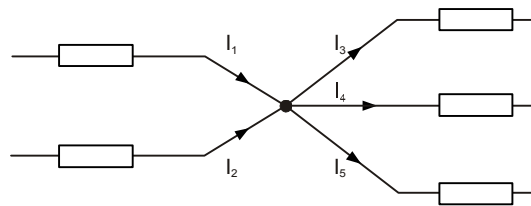
- a. *The First Law (Current Law).* This law states that at any junction in an electric circuit, the total current flowing towards that junction is equal to the total current away from the junction. For the circuit shown in Fig 5:

$$I_1 + I_2 = I_3 + I_4 + I_5$$

or

$$I_1 + I_2 - I_3 - I_4 - I_5 = 0$$

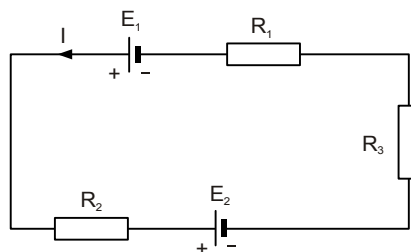
**14-1 Fig 5 Kirchhoff's First Law**



- b. *The Second Law (Voltage Law).* The second law states that in any closed loop in a network, the algebraic sum of the voltage drops (ie products of current and resistance) taken around the loop is equal to the resultant emf in that loop. For the circuit shown in Fig 6:

$$E_1 - E_2 = IR_1 + IR_2 + IR_3$$

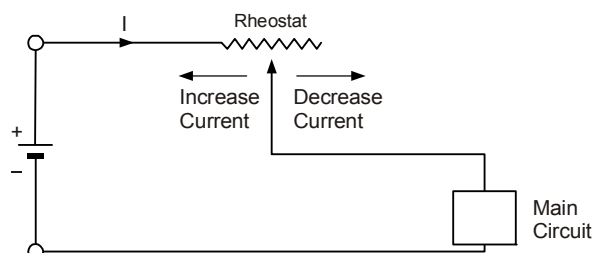
**14-1 Fig 6 Kirchhoff's Second Law**



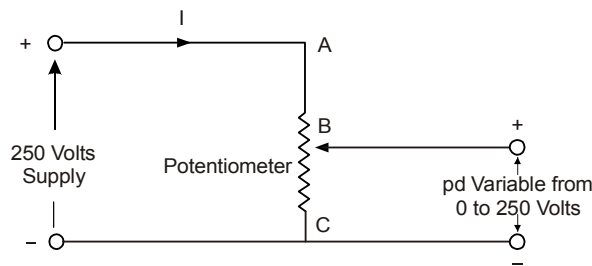
## Rheostats and Potentiometers

17. A rheostat is a variable resistor which is placed in series with a circuit, as shown in Fig 7, in order to control the current.

**14-1 Fig 7 A Rheostat**



18. A potentiometer is a tapped resistor which is connected across a pd source and used to control the voltage applied to the main circuit (see Fig 8).

**14-1 Fig 8 A Potentiometer**

19. Since the potentiometer is connected directly across the supply, its resistance must be large enough to prevent a heavy current being taken from the supply. If the load on the potentiometer takes an appreciable amount of current then the ratio between the two pds (A-B and B-C) will not be quite the same as the ratio of the two resistances of the potentiometer.

### Energy and Power

20. When an electric current flows in a conductor, energy is expended, ie work is done. The amount of energy expended is given by the product of the quantity of the charge moved and the difference in potential through which it moves. The unit of energy is the joule, therefore:

$$\begin{aligned} W \text{ (joules)} &= Q \text{ (coulombs)} \times V \text{ (volts)} \\ &= V \text{ (volts)} \times I \text{ (current)} \times t \text{ (time)} \\ \text{Since } Q &= \text{current} \times \text{time} \end{aligned}$$

Power (P) is measured in watts and is the rate of expenditure of energy, or work done per second.

$$\begin{aligned} P \text{ (watts)} &= \frac{V \text{ (volts)} \times I \text{ (current)} \times t \text{ (time)}}{t \text{ (time)}} \\ &= VI \text{ or } I^2 R \text{ or } \frac{V^2}{R} \end{aligned}$$

21. A joule can therefore be called a watt-second, but for practical purposes the joule is far too small a unit to work with, and for this reason the Kilowatt-hour ( $3.6 \times 10^6$  joules) is used as the standard unit of energy.

### Heating Effect of a Current

22. The energy expended in a purely resistive conductor is converted into heat. The unit of heat energy is the calorie, which is defined as the amount of heat required to raise the temperature of one cubic centimetre of water through one degree centigrade. One calorie of heat is produced in a conductor for every 4.2 joules of electrical energy expended. The total heat produced in any conductor is given by:

$$\frac{V \text{ (voltage)} \times I \text{ (current)} \times t \text{ (time)}}{4.2} \text{ calories}$$

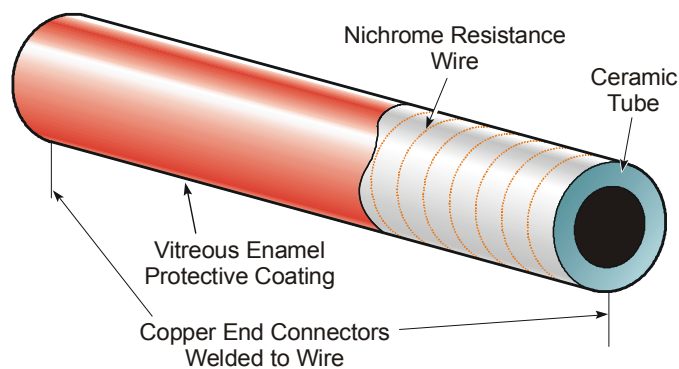
### The Maximum Power Transfer Theorem

23. The power transferred from a supply source to a load is at a maximum when the resistance of the load is equal to the internal resistance of the source. Referring to the circuit shown in Fig 2, maximum power is delivered when the load is equal to 1 ohm. This theorem is particularly important in the design of electronic equipment where maximum power transference is needed between circuit stages.

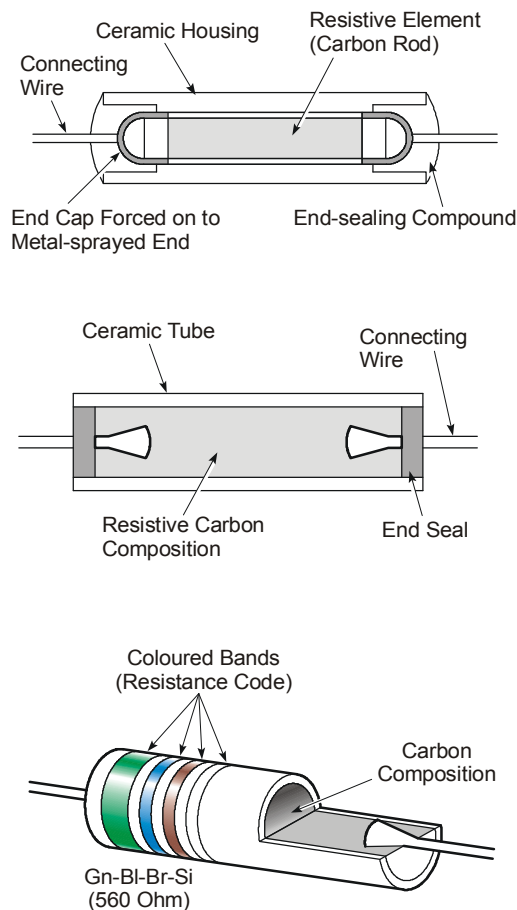
## Practical Resistors

24. There are several different types of resistor in common use, however, the three main types are carbon, metal film and wire-wound. General purpose resistors are more often of the carbon type; they are inexpensive and perform reasonably well in circuits where the design requirements are not too critical. Where higher accuracy components are called for, carbon or metal film resistors (low power) and wire-wound resistors (high power) are used. Resistors can overheat when overloaded and those installed in aircraft are therefore housed within a ceramic coating or tube. Fig 9 shows the construction of a typical high power wire-wound resistor and Fig 10 shows the composition of two types of carbon resistor, and a colour coded exterior.

**14-1 Fig 9 Construction of a Wire-wound Resistor**



**14-1 Fig 10 Two Types of Carbon Resistor**





## Preferred Values of Resistance

25. It is desirable to keep the number of resistance values to be manufactured or held in stock to a reasonable figure. In addition, the final value of a carbon composition resistor is rarely exactly the value that was originally intended. Because of these factors a logarithmic series of preferred resistor values has been chosen. For the  $\pm 20\%$  tolerance range, the ohmic values are 10, 12, 22, 33, 47, and 68, while the values for the  $\pm 10\%$  range are 10, 12, 15, 18, 22, 27, 33, 39, 47, 56, 68, and 82. Resistors manufactured to these values (or these values multiplied by factors of 10) are termed 'preferred value resistors'.

## The Resistor Colour Code

26. In order to identify the value of resistors and their tolerance range, methods of exterior markings have been devised. One such method of identification is by means of coloured bands. The bands are coloured according to a standard code; each colour representing a digit as indicated in Table 1.

**Table 1 Colour Code for Fixed Resistors**

Colour	Significant Figures	Multiplier	Tolerance
Silver	-	$10^{-2}$	$\pm 10\%$
Gold	-	$10^{-1}$	$\pm 5\%$
Black	0	1	-
Brown	1	10	$\pm 1\%$
Red	2	$10^2$	$\pm 2\%$
Orange	3	$10^3$	-
Yellow	4	$10^4$	-
Green	5	$10^5$	$\pm 0.5\%$
Blue	6	$10^6$	$\pm 0.25\%$
Violet	7	$10^7$	$\pm 0.1\%$
Grey	8	$10^8$	-
White	9	$10^9$	-
None	-	-	$\pm 20\%$

27. The coloured bands are painted on the body of the resistor usually biased towards one end. The colour of the band nearest the end indicates the first number. For easy recognition, on some resistors this band may be wider than the others, especially if there is likely to be any doubt over choice of ends. The colour of the second band gives the second number and the third band provides the multiplier (decimal place). The fourth band indicates the resistor's tolerance and will most often be either silver (10% tolerance) or gold (5% tolerance). Neither of these colours is used for significant figures, so it will be clear that counting of the value should begin at the other end. If there is no fourth band, the tolerance is 20%. Fig 11 shows a selection of colour-coded resistors with their values decoded together with a wire-wound example which has its value printed on the body.

**14-1 Fig 11 Examples of Resistors with their Values**



a Brown-Red-Black-Silver =  $12\Omega \pm 10\%$



b Red-Red-Black-Silver =  $22\Omega \pm 10\%$



c Red-Red-Brown-Silver =  $220\Omega \pm 10\%$



d Wire-wound Resistor Inscribed  $18\text{ M}\Omega$

## CHAPTER 2 - MAGNETISM

Contents	Page
Introduction .....	2
Magnetic Fields .....	2
Flux and Magnetic Circuits .....	4
Magnetic Flux by an Electric Current .....	4
Magnetomotive Force and Field Strength .....	6
Permeability .....	6
Reluctance .....	7
Series and Parallel Magnetic Circuits .....	8
Ferromagnetism .....	9
Magnetic Hysteresis .....	9
TERRESTRIAL MAGNETISM .....	10
Introduction .....	10
Magnetic Field Components .....	11
Magnetic Variations and Anomalies .....	11

### Table of Figures

14-2 Fig 1 The Magnetic Field Around a Bar Magnet .....	3
14-2 Fig 2 Modified Magnetic Field Patterns .....	3
14-2 Fig 3 The Magnetic Circuit .....	4
14-2 Fig 4 Magnetic Fields produced by Straight and Looped Wires .....	4
14-2 Fig 5 Current Flow Representation .....	5
14-2 Fig 6 The Magnetic Field of a Solenoid and the Grip Rule .....	5
14-2 Fig 7 The Polarity Rule .....	6
14-2 Fig 8 A Simple Magnetic Circuit .....	6
14-2 Fig 9 Comparison between a Magnetic and an Electric Circuit .....	7
14-2 Fig 10 Series Magnetic Circuit .....	8
14-2 Fig 11 A Parallel Magnetic Circuit .....	8
14-2 Fig 12 Magnetization Curves .....	9
14-2 Fig 13 Magnetic Hysteresis Loop .....	9
14-2 Fig 14 Hysteresis Loops for Hard and Soft Materials .....	10
14-2 Fig 15 The Earth's Magnetic Field .....	10

## Introduction

1. Around 3000 BC an iron ore known as lodestone was discovered in the Chinese desert. The ancient Chinese found that this ore possessed some unusual properties namely:

- a. The attraction or repulsion of other pieces of lodestone.
- b. The attraction of pieces of soft iron.
- c. Orientation in the direction of the Earth's North and South poles when suspended freely.

The ability of a piece of lodestone to align itself with the Earth's North and South poles led to the piece ends being called 'poles', and were labelled North (seeking) and South (seeking) poles. Such a piece of lodestone was called a magnet and the force which it possessed referred to as magnetism.

2. Magnetism was thought of as a force in its own right until Ampere discovered that a small coil carrying an electric current behaves like a magnet. It was suggested that electric and magnetic forces are both manifestations of the same electromagnetic force. Ampere's theory, which gave a natural explanation of the fact that no isolated magnetic pole had ever been observed, is essentially similar to modern day atomic theory. Electric charges can either be positive or negative, and the forces act so as to repel like charges and attract unlike charges. A moving electric charge produces a magnetic field. The electric current flowing through a coil of wire produces a controllable magnetic field, while in the case of a normal bar magnet the moving charges are the electrons circulating inside the atoms.

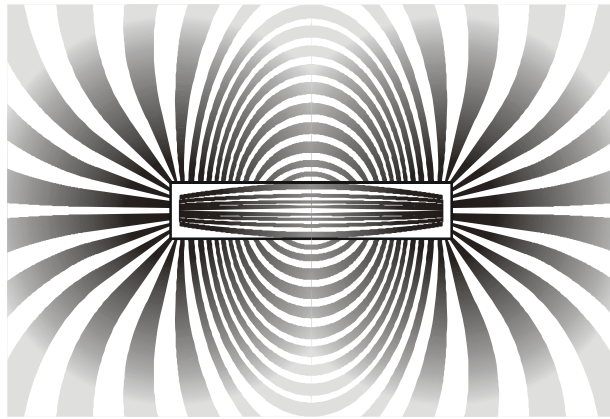
3. All materials exhibit magnetic properties, the degree to which they are exhibited depends on the distribution of electrons in the outer shells of the material's atoms. Materials can be placed in one of the following main magnetic categories:

- a. **Diamagnets.** Materials which oppose an applied magnetic field. Examples of such materials are copper, bismuth and hydrogen.
- b. **Paramagnets.** Materials which slightly aid an applied magnetic field.
- c. **Ferromagnets.** Materials which greatly aid an applied magnetic field. Examples of such materials are iron, steel, nickel and cobalt. Ferromagnetism will be discussed in more detail later in this chapter.

## Magnetic Fields

4. A magnet affects the space around it in such a way that other materials placed in this space experience forces. The space in which this occurs is known as a magnetic field, and its presence can be detected using iron filings or a compass needle. The magnetic field pattern, formed by iron filings around a bar magnet is shown in Fig 1. Michael Faraday referred to the apparent lines as lines of magnetic flux, and although flux does not exist as separate lines the concept of flux is useful in order to explain the effects of magnetism. Magnetic flux has the following properties:

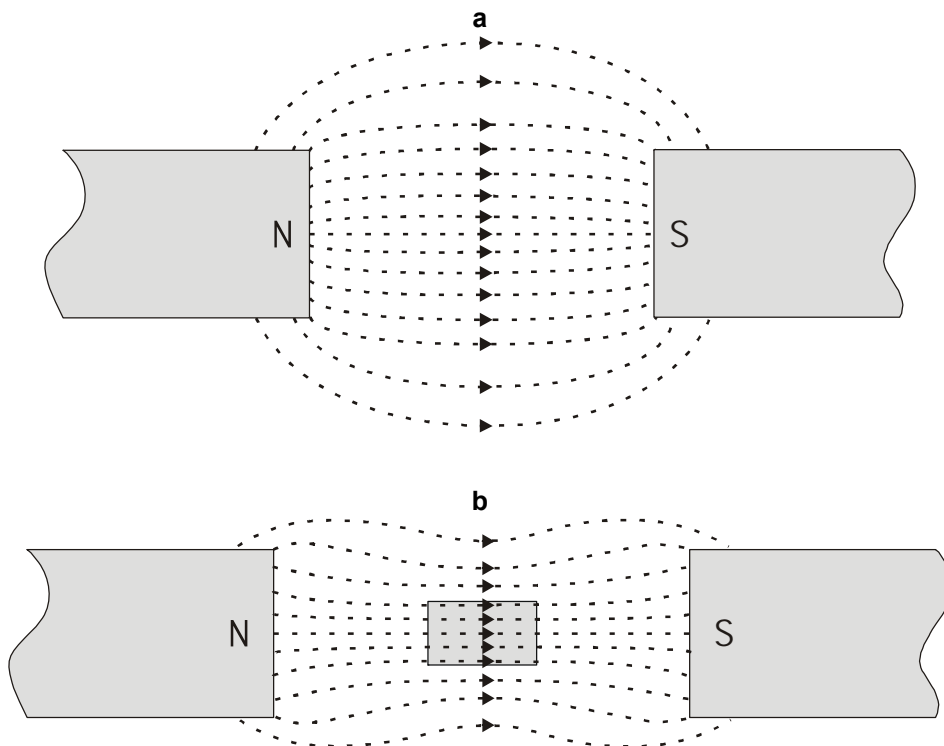
- a. The direction of the apparent lines, outside the magnet, is from the North to the South pole.
- b. Lines of flux form complete closed paths, they cannot end in space. Lines which do not close around the magnet join those of the Earth's magnetic field.
- c. Lines of magnetic flux never intersect each other, although may become extremely distorted.
- d. Where the flux is more intense, then the lines are closer together.

**14-2 Fig 1 The Magnetic Field Around a Bar Magnet**

5. Magnetic lines of force associated with a given magnet are modified in the presence of other magnets and magnetic materials. Fig 2 illustrates the effective fields produced by:

- a. Two adjacent, unlike poles
- b. The insertion of an iron bar in the space between the poles.

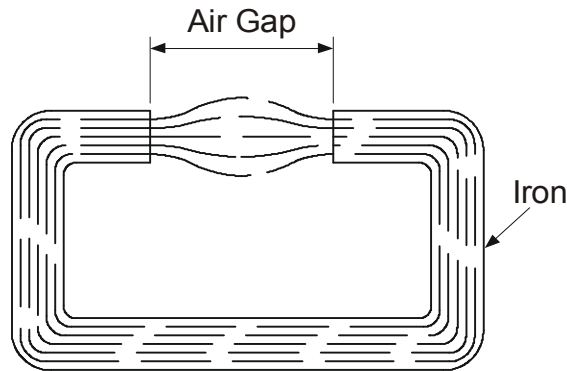
In the latter case, the flux lines appear to be concentrated in the vicinity of the iron. This is because the iron becomes magnetized by the external field and produces flux lines of its own, which reinforces the original field. This is referred to as induced magnetism.

**14-2 Fig 2 Modified Magnetic Field Patterns**

## Flux and Magnetic Circuits

6. Magnetic flux may be likened to an electric current, inasmuch as it only exists in circuits. The closed path in which the magnetic flux exists is known as a magnetic circuit. An example of a magnetic circuit is shown in Fig 3. The flux 'flows' through the iron ring and across the air gap, thus completing the magnetic circuit. The symbol for magnetic flux is  $\Phi$  (phi) and the unit is the Weber (Wb).

**14-2 Fig 3 The Magnetic Circuit**



7. The amount of Flux passing through a defined area, which is perpendicular to the direction of the flux, is referred to as the flux density, and is a far more useful quantity to work with than the total amount of flux. The symbol for flux density is  $B$  and the unit is the tesla (abbreviated as 'T'). Therefore:

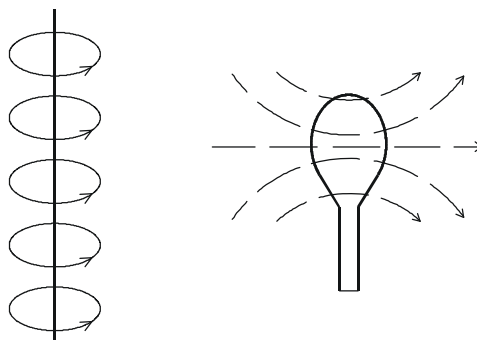
$$\text{Flux Density (B)} = \frac{\text{Total Magnetic Flux } (\Phi)}{\text{Area (A)}}$$

$$\text{and } 1 \text{ tesla} = \frac{1 \text{ weber}}{1 \text{ square metre}} \text{ or } 1 \text{ T} = 1 \text{ Wb/m}^2.$$

## Magnetic Flux by an Electric Current

8. When an electric current flows in a wire, a magnetic field is set up around the wire. This field may be represented by lines of force, with the strongest force being experienced near the wire. If the wire is bent to form a loop, the field will be strengthened at the centre of the loop (see Fig 4).

**14-2 Fig 4 Magnetic Fields produced by Straight and Looped Wires**



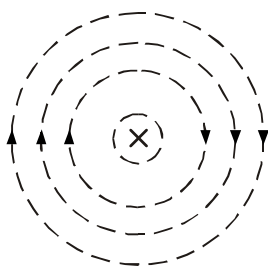
9. The direction of the magnetic field produced by a current carrying wire can be determined by applying Maxwell's corkscrew rule, which states:

If a normal right-hand threaded corkscrew travels along a conductor in the direction of the current, then the direction of rotation of the corkscrew is in the direction of the magnetic field.

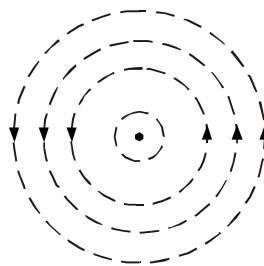
The convention that is used to show the direction of the current in a conductor, in diagrammatic form, is that the current flowing into the paper is shown by the flight of an arrow, while a current flowing outwards is shown by the arrow point. This is illustrated by Fig 5.

**14-2 Fig 5 Current Flow Representation**

**a Inward Current Flow**



**b Outward Current Flow**

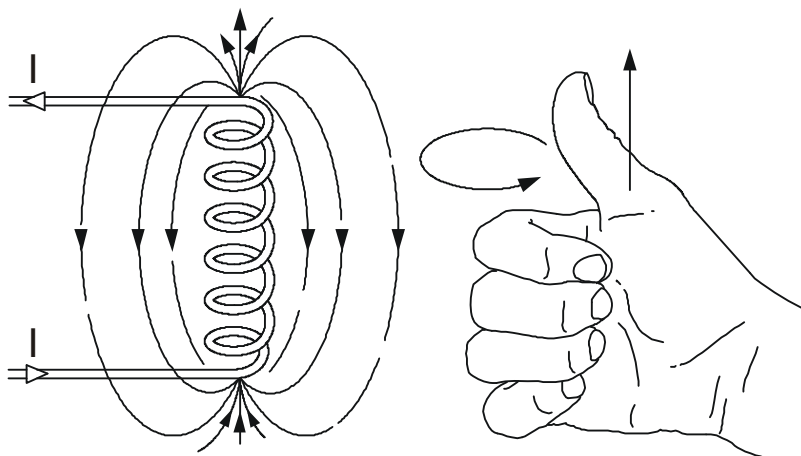


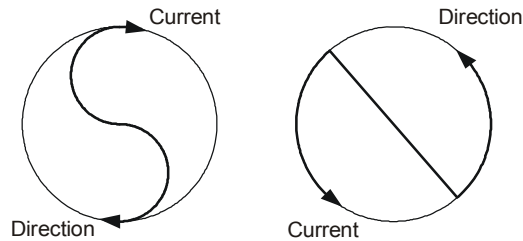
10. If a wire conductor is wound in the shape of a coil, then this is referred to as a solenoid. The direction of the magnetic field associated with a solenoid can be established using either the corkscrew rule or the more popular right hand grip rule. The grip rule states:

If a coil is gripped with the right hand, and with the fingers pointing in the direction of the current, then the thumb outstretched parallel to the axis of the solenoid, points in the direction of the magnetic field inside the solenoid (ie points in the direction of the North pole).

The magnetic field associated with a solenoid and the grip is shown in Fig 6. Fig 7 illustrates a simple method of determining the polarities of the solenoid ends.

**14-2 Fig 6 The Magnetic Field of a Solenoid and the Grip Rule**



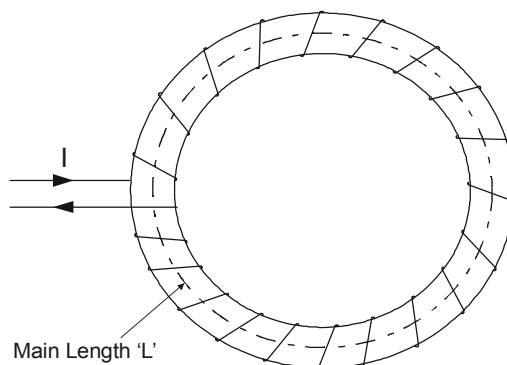
**14-2 Fig 7 The Polarity Rule****Magnetomotive Force and Field Strength**

11. In an electric circuit a current is established due to the existence of an electromotive force (emf). Similarly, in a magnetic circuit, a magnetic flux is established due to the existence of a magnetomotive force (mmf). The mmf is produced by the current in the coil and its magnitude depends on the number of turns of the coil ( $n$ ) and the current ( $I$ ). The unit of mmf is the ampere-turn (At), however, as the number of turns has no units, the alternative unit for mmf is the ampere (A). Magnetomotive force is represented by the symbol 'Fm'. Therefore:

$$\text{mmf} = F_m = nI$$

12. An alternative quantity to express the magnetic force produced by a current is magnetic field strength. This is the magnetomotive force per unit length of the magnetic circuit and is given by the symbol  $H$ . In the magnetic circuit shown in Fig 8, where a coil of  $n$  turns is uniformly wound around a metal ring, and  $L$  is the length of the magnetic circuit, the magnetic field strength is given by:

$$H = \frac{F_m}{L} = \frac{nI}{L} \text{ amperes per metre (A/m)}$$

**14-2 Fig 8 A Simple Magnetic Circuit****Permeability**

13. The magnetic flux density ( $B$ ) inside a current carrying coil is related to the magnetic field strength ( $H$ ), since one exists as a result of the other. The ratio of these two quantities is known as permeability and may be expressed as the ease with which a magnetic flux is set up within a magnetic circuit. Permeability is represented by the symbol  $\mu$  (mu) and the unit of measurement is the henry per metre (H/m). The henry is the unit of inductance and is considered in Volume 14, Chapter 3.

$$\text{Permeability } (\mu) = \frac{\text{magnetic flux density (B)}}{\text{magnetic field strength (H)}}$$



14. The magnitude of the permeability is dependent on the material placed in the centre of the coil. If the coil is placed in a vacuum the permeability is then referred to as the permeability of free space and is given the symbol  $\mu_0$ . This is a constant and has the value:

$$\mu_0 = 4\pi \times 10^{-7} \text{ H/m}$$

15. If a bar of magnetic material is placed in the centre of the coil then the flux density for the same magnetic field strength is greatly increased. Under these conditions the ratio of the flux density produced by the material to the flux density in a vacuum (or air) is called the relative permeability. This is a measure of the number of times the permeability of the material is greater than that of air. Relative permeability has the symbol  $\mu_r$ , and has no units, since it is a ratio of like quantities.

$$\text{Relative Permeability } (\mu_r) = \frac{\text{Flux density with magnetic core}}{\text{Flux density with air/vacuum core}}$$

The absolute permeability ( $\mu$ ) of the material is given by:

$$\mu = \mu_0 \mu_r = \frac{B}{H}$$

### Reluctance

16. The reluctance of a magnetic circuit is its opposition to the establishment of magnetic flux, and it may be linked, by analogy, to the resistance of an electrical circuit. Reluctance has the symbol 'S' and is the ratio of mmf ( $F_m$ ) to flux ( $\Phi$ ). Reluctance has the unit ampere per weber (A/Wb).

$$\text{Reluctance (S)} = \frac{\text{mmf (Fm)}}{\text{flux } (\Phi)}$$

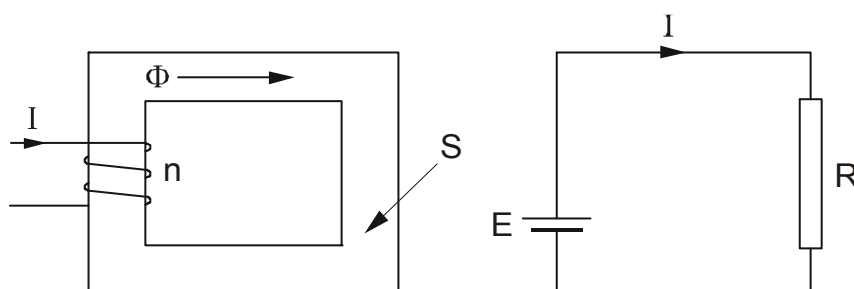
17. In Fig 9, a comparison is made between the magnetic circuit and the electric circuit. This analogy is a useful one when performing calculations on magnetic circuits, but it should be noted that, although flux may be compared to current, it does not flow. It simply exists as a result of the mmf. In the electric circuit, the emf (E) causes the current (I) to flow through the resistance (R), giving the relationship of:

$$E = IR$$

In the magnetic circuit, it is the mmf ( $F_m$ ) which causes the flux ( $\Phi$ ) to exist in the magnetic circuit of reluctance (S). The relationship is:

$$F_m = \Phi S$$

**14-2 Fig 9 Comparison between a Magnetic and an Electric Circuit**



### Series and Parallel Magnetic Circuits

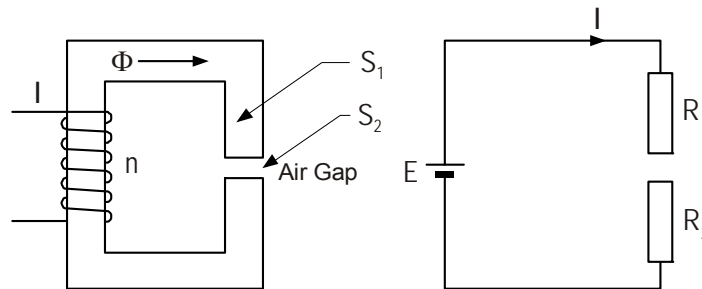
18. Fig 10 shows a simple series magnetic circuit consisting of a metal ring with an air gap. Neglecting losses, it can be assumed that the same flux ( $\Phi$ ) exists both in the iron ring and the air gap. The air gap has a large effect on the magnitude of the flux that exists in the circuit. If  $S_1$  is the reluctance of the metal ring and  $S_2$  the reluctance of the air gap, then the magneto-motive force required to maintain the flux is given by:

$$F_m = \Phi S_1 + \Phi S_2$$

This may be compared with the series electric circuit shown in the same figure.

$$E = IR_1 + IR_2$$

**14-2 Fig 10 Series Magnetic Circuit**



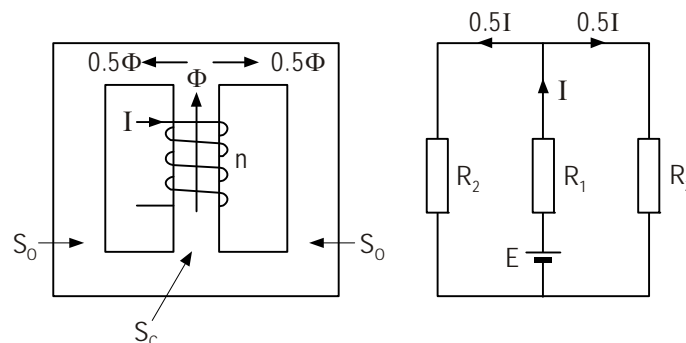
19. Fig 11 shows a parallel magnetic circuit such as is used in a certain type of transformer. The flux ( $\Phi$ ) which exists in the centre limb splits evenly between the two outer limbs so that each carries half of the flux,  $0.5\Phi$ . If  $S_c$  is the reluctance of the centre limb and  $S_o$  the reluctance of each outer limb, then, by considering only one of the magnetic circuits, the mmf is given by:

$$F_m = \Phi S_c + 0.5\Phi S_o$$

This equation can be compared with that of the parallel electrical circuit shown in the same figure. This time applying Kirchhoff's law and considering one loop:

$$E = IR_1 + 0.5IR_2$$

**14-2 Fig 11 A Parallel Magnetic Circuit**



## Ferromagnetism

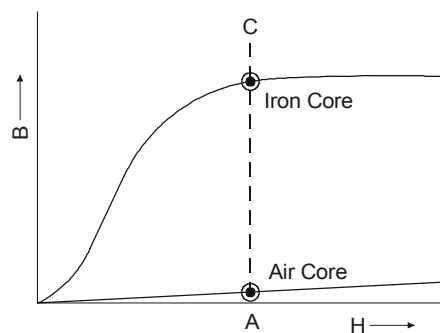
20. As stated earlier in this chapter, certain materials when magnetized produce their own magnetic field which greatly aids the applied field. These materials are known as ferromagnetic and have the ability to increase the magnetic flux density by as much as 1,000 times.

21. Ferromagnetic materials are classified as either hard or soft. Hard materials, such as cobalt steel and certain alloys, have a low relative permeability and are difficult to magnetize and demagnetize. On the other hand, soft materials, such as silicon and soft iron, have high relative permeability and are easily magnetized and demagnetized. Permanent magnets are made from materials with hard ferromagnetic properties, and areas of equipment where magnetic influences need to be minimized are usually protected using soft materials. This latter ferromagnetic feature is known as screening.

## Magnetic Hysteresis

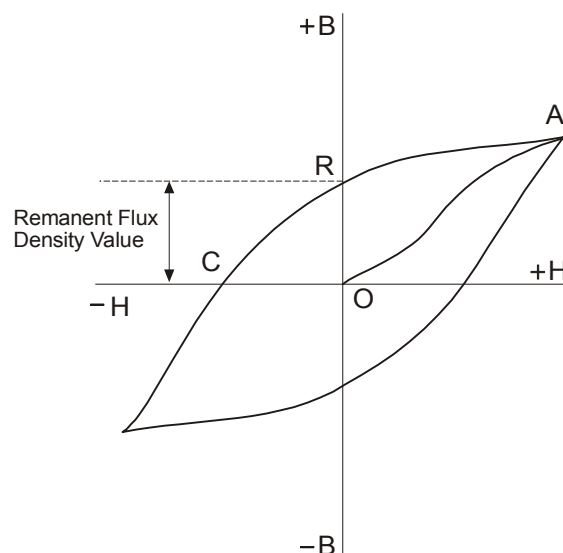
22. The usefulness of a piece of ferromagnetic material depends on its particular magnetic properties, and these can be recorded on a magnetization curve. This is a graph of flux density ( $B$ ) plotted against magnetic field strength ( $H$ ). A typical magnetization curve for a ferromagnetic material is shown in Fig 12; the curve for air is also shown for comparison.

14-2 Fig 12 Magnetization Curves



23. It is evident that the slope of the magnetization curve for the ferromagnet is not constant and, at the point C, increasing the magnetizing force has little effect on flux density; this is known as saturation. If the magnetizing force is now reduced to zero the material will retain some of its magnetization, which is referred to as the remanent flux density, or remanence. It is represented by OR in Fig 13.

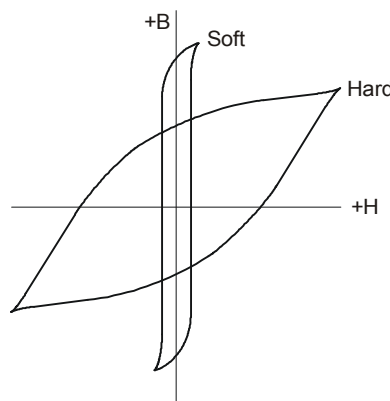
14-2 Fig 13 Magnetic Hysteresis Loop



24. If the magnetic field strength is reversed ( $-H$ ), then the flux density will eventually reach zero. This field, represented by  $OC$ , is known as the coercive force or coercivity of the material. Increasing the magnetic field strength still further, in the negative sense, magnetizes the material in the opposite direction.

25. When the magnetic field strength is increased and decreased, first in one direction and then in the other, the plotted values of  $B$  and  $H$  describe a loop which is referred to as the hysteresis loop of the material. The area under the curve is proportional to the work carried out in order to complete the cycle and as such represents losses. Fig 14 illustrates the differences in the hysteresis loops for hard and soft ferromagnetic materials.

**14-2 Fig 14 Hysteresis Loops for Hard and Soft Materials**

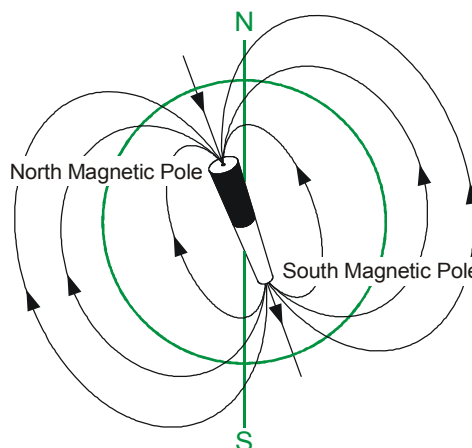


## TERRESTRIAL MAGNETISM

### Introduction

26. The centre of the Earth is made up of a core of magnetic material whose magnetic field of influence extends beyond the surface of the planet. The core is aligned in such a way that its magnetic poles correspond approximately to the Earth's geographical poles. In essence, the Earth can be regarded as being similar to a large bar magnet with its lines of force running from geographical South to geographical North. This field is depicted in Fig 15.

**14-2 Fig 15 The Earth's Magnetic Field**



**Magnetic Field Components**

27. At any point on the Earth's surface, the magnetic field can be resolved into its vertical and horizontal components. The horizontal component is used extensively for navigation and direction finding by means of a compass. The angle formed between the horizontal and the actual force line is referred to as the angle of dip. The angle of dip varies between  $0^\circ$  at the magnetic equator and  $90^\circ$  at the magnetic poles.

**Magnetic Variations and Anomalies**

28. The Earth's North and South magnetic poles are not coincident with the geographical poles; there is an angle between the true North and the magnetic North. The angle between them, known as variation, is a variable quantity and describes a circle around the geographical poles roughly every 1,000 years. In England, the variation is seen as a swing from about  $27^\circ$  W to  $27^\circ$  E. This long-term movement in magnetic variation is modulated by daily changes (Diurnal), annual fluctuations (opposite directions in northern and southern hemispheres) and changes which coincide with 11 year sunspot cycle, (Periodic).

29. The Earth's magnetic field can also be disturbed by metallic objects placed within its field. In the same way that an iron rod distorts the magnetic field of a bar magnet due to induced magnetism, underground cables and shipwrecks distort the Earth's magnetic field. These disturbances, or anomalies, can be detected using an instrument known as a magnetometer. The magnetometer principle is dealt with in Volume 14, Chapter 5.

## CHAPTER 3 - ELECTROMAGNETIC INDUCTION AND INDUCTANCE

Contents	Page
Introduction .....	2
Magnetism as a Source of Electricity .....	2
Electromagnetic Induction .....	3
Inductive DC Series Circuits .....	4
Inductive DC Parallel Circuits .....	6
Inductive-Resistive Circuits, Time Constants .....	6

### Table of Figures

14-3 Fig 1 A Simple Generator .....	2
14-3 Fig 2 Mutual Induction .....	3
14-3 Fig 3 Self-Induction .....	4
14-3 Fig 4 Inductors in Series .....	5
14-3 Fig 5 Growth and Decay of Current in an Inductive DC Circuit .....	5
14-3 Fig 6 Inductors in Parallel .....	6

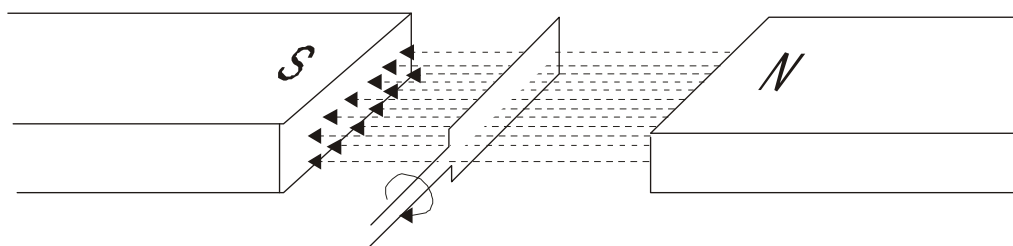
## Introduction

1. The phenomenon of electromagnetic induction forms the basis of many electrical machines, such as the dc motor, dc generator, induction motor, synchronous motor and the transformer. Since these are the most frequently used machines in electrical engineering, it is important that the laws of magnetic induction and inductance are clearly understood.

## Magnetism as a Source of Electricity

2. When a magnetic field is moved across a conductor, or conversely when a conductor is moved across a stationary magnetic field, an emf is induced in the conductor. This can be demonstrated by moving a magnet into and out of a coil connected to a sensitive ammeter (a device to measure current), or by rotating a loop connected to an ammeter within a magnetic field so as to cut across the field (Fig 1). In both cases relative movement of the conductor and the magnetic field will produce a reading on the ammeter, showing that a current has been created as a result of the induced emf.

14-3 Fig 1 A Simple Generator



3. It can be demonstrated that:

- a. The emf is present only whilst there is relative motion between the conductor and the magnetic field.
- b. The greater the relative motion, the greater the emf. Thus the rate at which the flux is changing relative to the conductor determines the magnitude of the induced emf.
- c. The stronger the magnetic field, the greater the induced emf.
- d. The relative motion must be such that the conductor cuts across the magnetic flux.
- e. Reversing the relative motion will reverse the direction of the induced emf.

4. To obtain a large voltage, a coil with a large number of turns must be used, and the magnetic field must be a strong one with a continuous high relative motion. One way to obtain rapid continuous relative motion is to rotate a loop in the magnetic field, as in Fig 1, which represents a simple electricity generator. In practice many turns are used in place of a single loop.

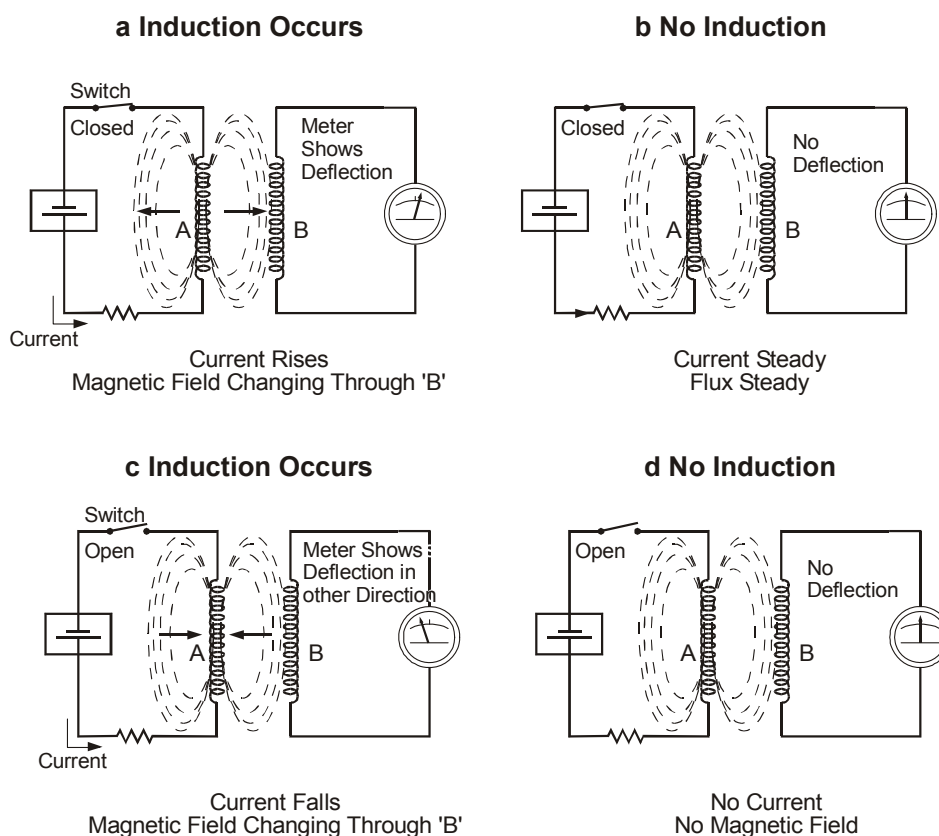
## Electromagnetic Induction

5. Two laws state the theory of electro-magnetic induction very concisely:

- a. **Faraday's Law.** When the magnetic flux through a circuit is changing an induced emf is set up in that circuit, and its magnitude is proportional to the rate of change of flux.
- b. **Lenz's Law.** The direction of an induced emf is such that its effect tends to oppose the change producing it.

6. **Mutual Induction.** Consider the two coils, A and B, connected as shown in Fig 2. When no current is flowing there is no magnetic field, but when the switch is closed a magnetic field expands outwards from the primary coil (A) and cuts the secondary coil (B). As there is a change of flux through the secondary coil an emf is induced in it and the meter shows a deflection. When the field stops growing there is no longer a change of flux through the secondary and no emf is induced. When the switch is opened the primary field collapses and contracts, thus flux changes in the secondary and an emf is induced again, but in the opposite direction. It is important to note that when the DC supply is steady no mutual induction occurs, but when the supply is AC induction occurs continuously.

14-3 Fig 2 Mutual Induction

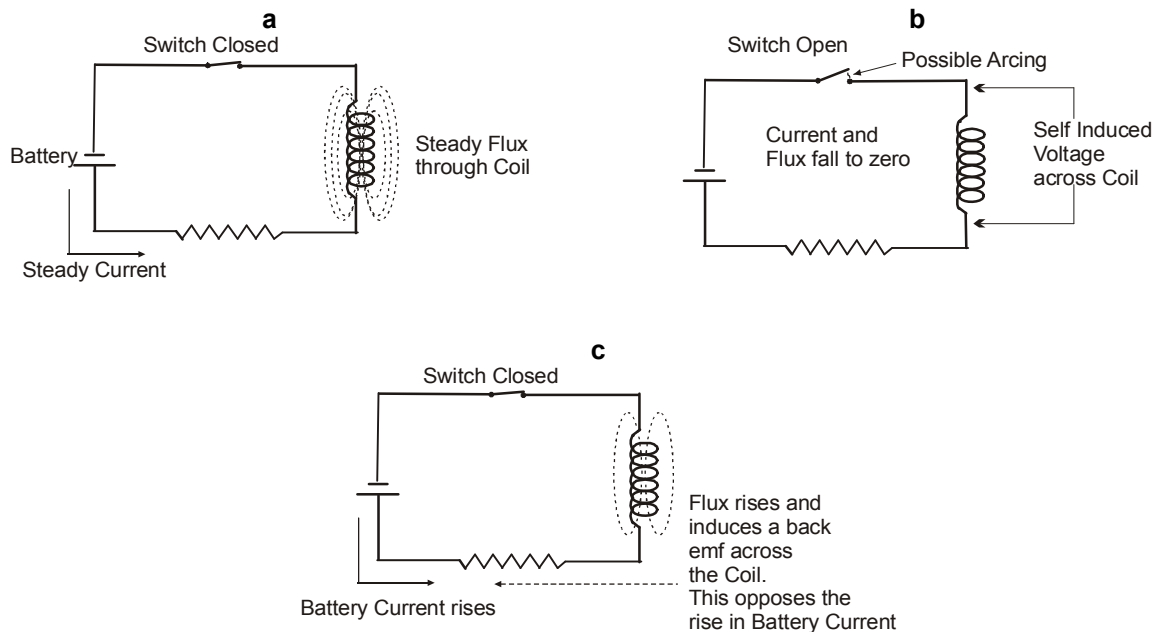


7. **Mutual Inductance (M).** The size of the induced emf depends partly on the rate at which the flux is changing. It also depends on various factors associated with the coils, these being the ratio of the number of turns of the primary and secondary coils, the relative positions of the coils, and the permeability of the medium through which the flux travels; together these are known as the mutual inductance of the coils (M).



8. **Self-Induction.** In the circuit shown at Fig 3 a magnetic field is established around a coil. When the switch is opened the field will start to collapse as the current reduces. There is therefore a change of flux through the coil, and an emf will be induced in it. The induced emf opposes the inducing emf (Lenz's Law), and thus is called the back emf. No back emf is created by a steady DC supply except at switch on and switch off, whereas an AC supply causes self-induction continuously.

14-3 Fig 3 Self-Induction



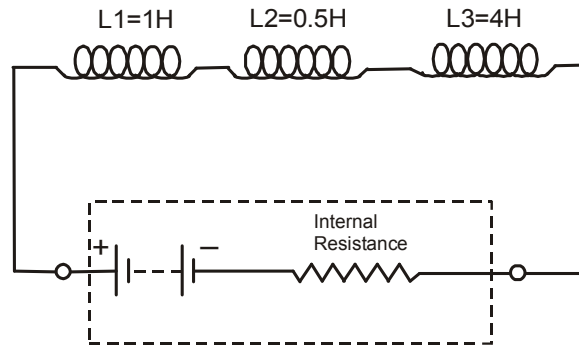
9. **Self-Inductance (L).** Any circuit that has a voltage induced in it by a change of current through the circuit itself has self-inductance. The property of self-inductance is to oppose any change of current by the creation of a back emf. The size of the back emf depends partly upon the rate of change of current. If the circuit includes a coil, the size of the back emf depends also on the number of turns in the coil and the permeability of the medium around the coil, these factors associated with the coil being termed the self-inductance (L).

10. **Units of Inductance.** Both mutual and self-inductance are measured in the same units - the henry (H). The mutual inductance of two circuits is 1 henry when a current changing uniformly at the rate of 1 ampere per second in one circuit produces a mutually induced emf of 1 volt in the other circuit. The self-inductance of a closed circuit is 1 henry when an emf of one volt is produced when the current in the circuit is changing uniformly at the rate of one ampere per second.

11. **Inductors.** When a coil is used specifically to provide inductance and thus oppose any change in current in the circuit it is called an inductor or choke. Typical inductance values of inductors in practical use range from about 100 H down to a few microhenrys according to the application.

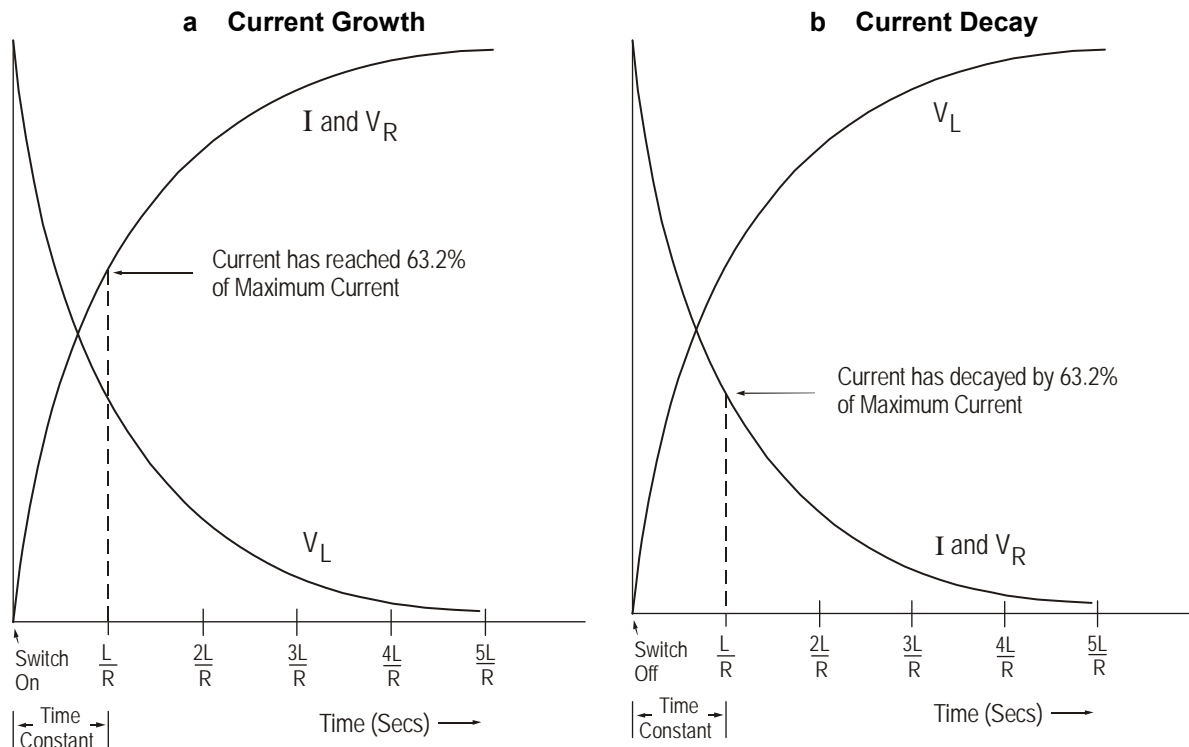
### Inductive DC Series Circuits

12. When several inductors are connected in series ( $L_1$ ,  $L_2$  and  $L_3$  in Fig 4) circuit inductance is found by summing the individual values, provided that there is no mutual inductance between the coils.

**14-3 Fig 4 Inductors in Series**

$$\begin{aligned}
 L &= L1 + L2 + L3 \\
 &= 1 + 0.5 + 4 \\
 &= 5.5 \text{ henrys}
 \end{aligned}$$

13. It was stated earlier in this chapter that the back emf created by the build-up or collapse of the magnetic field around an inductor opposes the change in current. A plot of circuit current against time takes the form of an exponential curve, rising for current increase or switch-on (Fig 5a), and falling with current decrease or switch-off (Fig 5b). It is impossible to create a purely inductive circuit because the conductor must have some resistance, and the resistive voltage,  $V_R$ , rises and falls as the circuit current rises and falls (Fig 5). On the other hand the voltage,  $V_L$ , across the inductor, falls as the current rises and rises as the current falls (Fig 5). The time taken for the current to reach particular values is discussed in para 15.

**14-3 Fig 5 Growth and Decay of Current in an Inductive DC Circuit**

### Inductive DC Parallel Circuits

14. When inductors are connected in parallel (Fig 6), providing alternative paths for the current flow, the total inductance of the circuit, provided that the coils are completely isolated from one another such that there is no mutual inductance, is given by:

$$\frac{1}{L} = \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \dots \frac{1}{L_n}$$

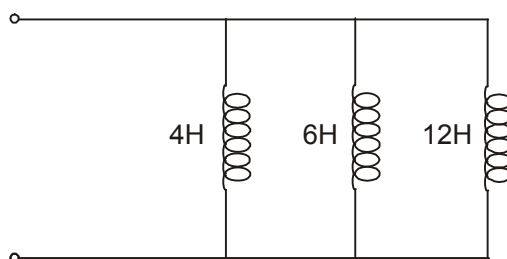
Applying this formula to the example in Fig 6:

$$\frac{1}{L} = \frac{1}{4} + \frac{1}{6} + \frac{1}{12}$$

$$= \frac{6}{12}$$

$$\therefore L = 2\text{H}$$

**14-3 Fig 6 Inductors in Parallel**



### Inductive-Resistive Circuits, Time Constants

15. The exponential rise and fall of current in an inductive-resistive DC circuit was discussed in para 13. In theory the current in an inductive circuit can never reach a maximum or fall completely to zero. However, for all practical purposes the growth or decay of current is complete in  $5\frac{L}{R}$  seconds,

where L is the inductance in henrys and R the resistance in ohms.  $\frac{L}{R}$  seconds is known as the time

constant of an inductive-resistive circuit. The time constant is defined as the time taken for the current through an inductive circuit to rise to 63.2% of its maximum value when connected to a supply, or to fall by 63.2% of its maximum value when disconnected from a supply. The time constant is indicated in Figs 5a and 5b.

## CHAPTER 4 - ELECTROSTATICS AND CAPACITANCE

Contents	Page
Introduction .....	2
ELECTROSTATICS .....	2
General .....	2
Coulomb's Law .....	2
The Electric Field .....	2
CAPACITANCE .....	4
The Simple Capacitor .....	4
Factors Affecting Capacitance .....	5
Safe Working Voltage .....	5
Factors Determining the Construction of Capacitors .....	6
Fixed Capacitors .....	6
Variable and Trimmer Capacitors .....	8
Energy Stored in a Charged Capacitor .....	9
Capacitive DC Series Circuits .....	9
Capacitive DC Parallel Circuits .....	10
Capacitive-Resistive Circuits, Time Constants .....	11
The Oscillatory Circuit .....	11

### Table of Figures

14-4 Fig 1 Typical Electric Field Pattern .....	3
14-4 Fig 2 The Electric Field Between Two Plates .....	3
14-4 Fig 3 Paper Capacitor .....	6
14-4 Fig 4 Mica Capacitor .....	7
14-4 Fig 5 Ceramic Capacitor .....	7
14-4 Fig 6 Electrolytic Capacitor .....	7
14-4 Fig 7 Variable and Trimmer Capacitors .....	8
14-4 Fig 8 Capacitors in Series .....	10
14-4 Fig 9 Voltage and Current during Charges in a Capacitive DC Circuit .....	10
14-4 Fig 10 Capacitors in Parallel .....	11
14-4 Fig 11 The Oscillatory Circuit .....	11
14-4 Fig 12 Oscillatory Damping due to Circuit Resistance .....	12

### Tables

Table 1 Dielectric Constants .....	5
Table 2 Capacitor Summary .....	9

## Introduction

1. This chapter opens with a brief mention of electrostatics, which, as the name implies, is primarily the science of electric charges at rest. It then goes on to discuss a related phenomenon known as 'capacitance', and describes the construction and properties of devices used to introduce capacitance into the circuit ('capacitors').

## ELECTROSTATICS

### General

2. Electrification by friction is a common phenomenon: a comb passed through dry hair attracts the individual hairs, which then tend to stand on end, repelling one another. If a glass rod is rubbed with a piece of silk, the silk is attracted towards the rod. In this case, the silk removes electrons from the glass which is thus left with a positive charge; the electrons acquired by the silk give it an equal negative charge. If two freely suspended glass rods are treated in this way, and then brought near to each other, a mutual repulsion is evident. From these facts, the first law of electrostatics can be stated: like charges repel each other; unlike charges attract.

### Coulomb's Law

3. It can be shown that the size of the force of attraction (or repulsion) is greater between large charges than between small charges, and is greater when the charges are close together than when they are more distant. Coulomb's law states that the force ( $F$ ) between two quantities of electricity ( $Q_1$  and  $Q_2$ ), placed a distance apart ( $d$ ), is proportional to the product  $Q_1Q_2$  and inversely proportional to  $d^2$ , ie  $F \propto Q_1Q_2/d^2$ .

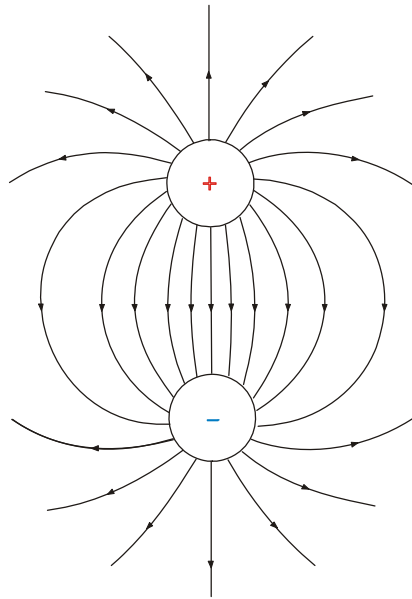
### The Electric Field

4. If an insulated metal sphere ( $A$ ) is positively charged, it will repel another positively charged body ( $B$ ) brought near to it; the nearer the two approach, the greater the repulsion. Work has to be done to overcome this repulsion; if the force drawing the bodies together is removed, the repulsion will force  $B$  away towards its original position (assuming  $A$  to be fixed). The system possesses potential energy, and a potential is said to exist at any point in the vicinity of  $A$ ; the nearer to  $A$ , the higher the potential. A positive charge placed in the vicinity of  $A$  will tend to move away, ie from a position of higher to lower potential. The change of potential per unit distance is known as potential gradient.

5. The region around a charged body is described as an electric field. Electric lines of force (akin to lines of magnetic flux) are used to show the distribution of an electric field, each line showing the direction in which a free positive charge would tend to move if placed in that field. Fig 1 shows a typical field pattern of two oppositely charged spheres. The unit of electric flux, or charge ( $Q$ ), is the coulomb, and the flux density ( $D$ ) is the charge per unit area:

$$\text{Flux Density} = \frac{\text{Charge (coulombs)}}{\text{Area (square metres)}}$$

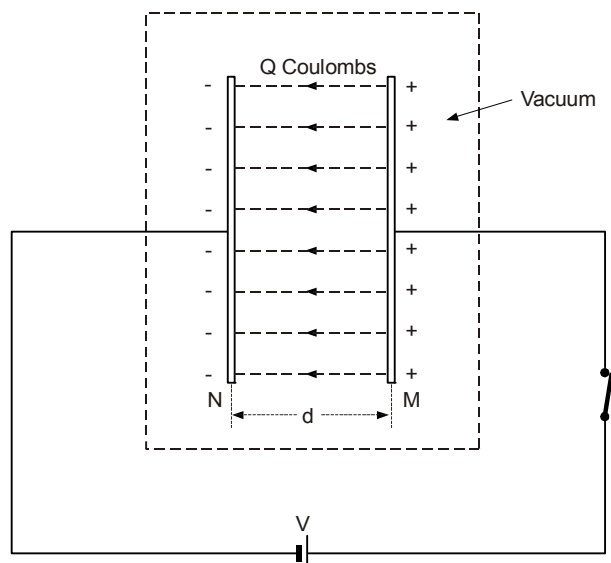
$$\text{or } D = \frac{Q}{A}$$

**14-4 Fig 1 Typical Electric Field Pattern**

6. **Electric Field Strength.** Two plates (M and N), separated by a gap (d), are connected to a battery as shown in Fig 2. Electrons begin to move from plate M towards the positive terminal of the battery, leaving the plate M positively charged. At the same time, electrons move from the negative plate of the battery to the plate N, which acquires a negative charge. A difference of potential will exist between the plates, and the space between them is an electric field. The intensity (E) of the electric field will depend on the PD and the distance between the plates:

$$\text{Intensity} = \frac{\text{Potential Difference (volts)}}{\text{Distance Apart (metres)}}$$

$$\text{or } E = \frac{V}{d}$$

**14-4 Fig 2 The Electric Field Between Two Plates**

7. **Permittivity of Free Space.** The ratio of flux density (D) to the electric field strength (E) in free space is termed the 'permittivity of free space' ( $\epsilon_0$ ). Thus:

$$\epsilon_0 = \frac{D}{E} = \frac{Q}{A} \div \frac{V}{d} = \frac{Qd}{VA}$$

$$\text{But, } \frac{Q}{V} = C \text{ (see para 11).}$$

$$\therefore \epsilon_0 = C \frac{d}{A} \quad \text{and } C = \epsilon_0 \frac{A}{d} \text{ farads}$$

The value of the constant  $\epsilon_0$  is  $8.85 \times 10^{-12}$  F/m. (F/m: Farad per metre)

## CAPACITANCE

### The Simple Capacitor

8. To charge a conductor negatively, electrons are added to it. Initially, the electrons move easily but, as the conductor becomes more and more negatively charged, so the electrons on the conductor repel those which try to add to their number. Eventually, the repelling force equals the charging force and no further movement of electrons takes place. The conductor is then said to be fully charged.

9. The charging process is due to the applied voltage. The charge can be increased by increasing the applied voltage, but the conductor will oppose the change because of the repulsion by the electrons it already possesses. Similar remarks apply when a conductor is positively charged.

10. It has already been shown that if, instead of a single conductor, a pair of plates, as in Fig 2, is considered, a difference of potential will exist between the plates if connected to a battery. The charge held by the combination of plates for a given applied voltage can be made greater by sandwiching a suitable insulator (a 'dielectric') between them.

11. For a given pair of plates, increasing the charge increases the voltage between the plates, but the ratio of charge to applied voltage remains constant. This ratio is known as the 'capacitance' (C) of the capacitor. The unit of capacitance is the farad; a capacitor having a capacitance of one farad if a charge of one coulomb (i.e. a charging current of one ampere flowing for one second) causes a charge of one volt in the potential difference between the plates, i.e.:

$$\text{Capacitance (farads)} = \frac{\text{Charge (coulombs)}}{\text{Potential Difference (volts)}}$$

Because the farad is a very large unit, capacitances are usually given in microfarads ( $\mu\text{F}$  - a millionth of a farad, ie  $10^{-6}$  F) or picofarads (pF - a million-millionths of a farad, ie  $10^{-12}$  F).

## Factors Affecting Capacitance

12. The capacitance of a capacitor depends on the following factors:

- a. **The Area of the Plates.** Capacitance increases with increase in area (A) of the opposed surfaces.
- b. **The Distance Apart of the Plates.** Capacitance increases as the distance (d) between the plates decreases (ie by using a thinner dielectric), because the field then becomes more concentrated.
- c. **The Material of the Dielectric.** Different dielectrics produce different capacitances in capacitors where other factors (plate area and distance between the plates) are equal. For example, substituting waxed paper for air increases the capacitance by a factor of about three. The ratio of the capacitance of a capacitor having a certain material as dielectric to the capacitance of the same capacitor having a vacuum as dielectric, is known as the 'dielectric constant', or 'relative permittivity' ( $\epsilon_r$ ) of the dielectric material.

13. It follows that the capacitance of a capacitor is given by:

$$C = \epsilon_0 \epsilon_r \frac{A}{d} \text{ (farads)}$$

A comparison of dielectric materials is given in Table 1. For practical purposes, air can be considered to have a dielectric constant of unity, as for free space.

**Table 1 Dielectric Constants**

Dielectric Material	Relative Permittivity
Dry air	1.00
Polypropylene	2.25
Polystyrene	2.50
Polycarbonate	2.80
Polyester	3.20
Impregnated paper	4 to 5
Mica	6.00
Aluminium oxide	7.50
Tantalum oxide	25.00
High-permittivity ceramic	10,000.00

## Safe Working Voltage

14. The safe working voltage is the maximum DC voltage that can safely be applied to a capacitor without causing the dielectric to break down; if this voltage is exceeded the electric field becomes strong enough to break down the insulation of the dielectric, a spark occurs, and the capacitor is usually ruined. Increasing the thickness of the dielectric enables the capacitor to withstand higher voltages, but, to compensate for the greater distance between the plates, a larger plate area is necessary to maintain the capacitance value. Hence, capacitors with high voltage ratings are larger



than similar types of capacitor of low rating. Note that a charged capacitor (especially one of large capacitance) can be dangerous; capacitors must be discharged before touching them.

### Factors Determining the Construction of Capacitors

15. Capacitors are used for a wide variety of purposes in electronics. In their construction, three factors associated with the dielectric are important:

- a. The dielectric constant, discussed in paras 12 and 13.
- b. The dielectric strength.
- c. Dielectric losses.

16. **The Dielectric Strength.** Dielectric strength is a measure of the PD required to break down the dielectric and is normally given as kilovolts per millimetre thickness. For a given material, the breakdown voltage will depend on:

- a. **Thickness of the Dielectric.** Although not directly proportional, the thicker the dielectric, the greater the breakdown voltage.
- b. **Temperature.** An increase in temperature of the dielectric produces a decrease in the breakdown voltage.
- c. **Frequency.** If the applied voltage is AC, the higher the frequency the lower the breakdown voltage due to the higher temperature produced.

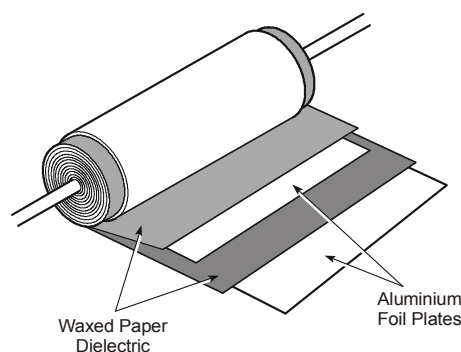
The safe working voltage of a capacitor is usually given as a DC voltage at a given temperature, eg "350V DC working, 71 °C". When used in an AC circuit, the rms voltage applied should not exceed 0.707 of the safe working DC voltage.

17. **Dielectric Losses.** A certain proportion of the energy applied to a capacitor on charging is not available as energy on discharging. Some energy is expended in the dielectric, being lost as heating, leakage current, and through other effects. The ratio of energy applied on charging to energy available on discharge is a measure of the capacitor efficiency.

### Fixed Capacitors

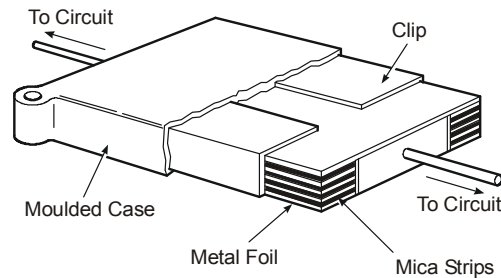
18. **Paper Type.** Capacitors with paper dielectrics are the commonest and cheapest fixed capacitors. The two plates are long strips of aluminium foil, separated by similar strips of waxed paper acting as the dielectric (Fig 3). This assembly is rolled up into a tube, which is then wax-sealed into an outer container of cardboard or metal.

14-4 Fig 3 Paper Capacitor



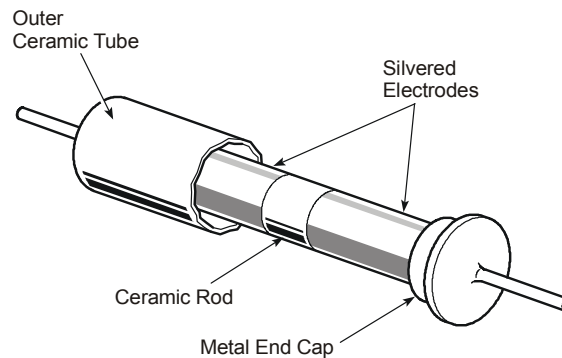
19. **Mica Type.** Mica Capacitors are high quality capacitors with several plates of metal foil interleaved with layers of mica acting as the dielectric (Fig 4). Alternate foils are joined together and connected to the two terminals. The assembly is placed inside a moulded plastic or metal case.

**14-4 Fig 4 Mica Capacitor**



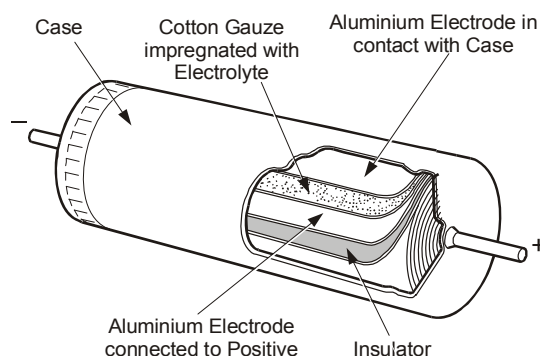
20. **Ceramic Type.** Ceramic capacitors can give very small values of capacitance. They use a ceramic disc or rod as the dielectric. Two films of silver, deposited on the ceramic, act as the plates (Fig 5).

**14-4 Fig 5 Ceramic Capacitor**



21. **Electrolytic Type.** For large values of capacitance, electrolytic capacitors are used, paper and mica types being too bulky. The large capacitance value is due to the very thin dielectric used. The electrodes are two long strips of aluminium foil. During manufacture, the dielectric film is formed on the positive electrode. The electrodes are separated by strips of cotton gauze impregnated with electrolyte. The whole assembly is tightly rolled (Fig 6) and mounted in a metal case. One of the aluminium electrodes provides a contact between the case and the electrolyte paste, which acts as the negative plate. The dielectric film can only be maintained if the capacitor has a DC component of voltage applied to it. The capacitor must be connected the right way round in the circuit, and any AC component must be less than the polarizing DC component.

**14-4 Fig 6 Electrolytic Capacitor**



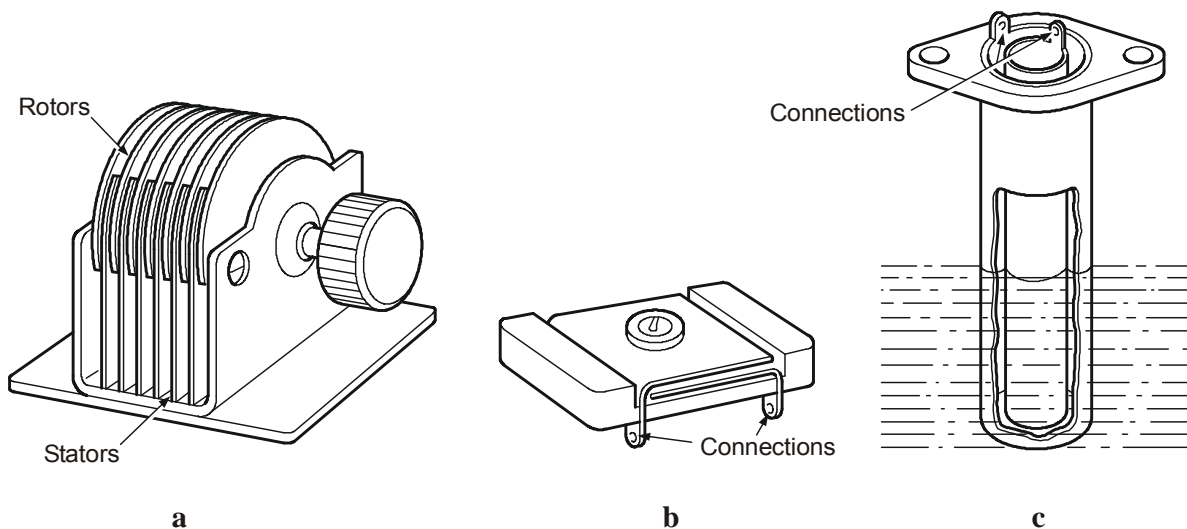
## Variable and Trimmer Capacitors

22. In electronic circuits, it is often necessary to be able to vary the capacitance. Variation of capacitance can be achieved by:

- a. **Varying the Effective Area of the Plates.** One set of plates (the rotors) can be moved into and out of mesh with the other set (the stators), so that the effective area of overlap is changed (Fig 7a).
- b. **Varying the Distance between the Plates.** The distance between the plates can be varied by a screw adjustment which squeezes the plates together (Fig 7b).
- c. **Changing the Dielectric.** If two (or more) metal plates are set at a fixed distance apart in a container, then, when the container is empty, the dielectric is air. If a fluid which has dielectric properties is pumped into the container, then the capacitance will change as the level of fluid changes (Fig 7c). This principle is applied in measuring the contents of storage tanks.

The types described in sub-para a and b are normally referred to as variable capacitors; a smaller version of that in sub-para a, and the type at sub-para b are used to make small variations of circuit capacitance, and are referred to as trimmer capacitors. The main properties and uses of the various forms of capacitor are summarized at Table 2.

14-4 Fig 7 Variable and Trimmer Capacitors



**Table 2 Capacitor Summary**

Type	Capacitance Values	DC Working Voltage (max)	Remarks
Paper	100 pF to 12 $\mu$ F	150 kV	Used in circuits where dielectric losses are unimportant. Cheap.
Mica	50 pF to 0.25 $\mu$ F	2 kV	Used in low-loss circuits. Expensive.
Ceramic	0.5 pF to 0.005 $\mu$ F	500 V	Used in low-loss precision circuits where micro-miniaturization is important or where temperature compensation is required.
Electrolytic	2 $\mu$ F to 32 $\mu$ F	600 V	Used where losses are not important - smoothing circuits.
	32 $\mu$ F to 3000 $\mu$ F	50 V	A polarizing DC voltage must be operative in the circuit to prevent reverse electrolytic action.
Variable	Variable from 50 pF to 500 pF	2 kV	Used for circuit tuning.
Trimmer	Variable from 2 pF to 50 pF	350 V	Used for circuit alignment.

**Energy Stored in a Charged Capacitor**

23. When an initially uncharged capacitor is charged at a constant rate of  $I$  amps for  $t$  seconds, the charge ( $Q$ ) equals  $(I \times t)$  coulombs. During the charge, the PD across the capacitor will have risen from zero to  $V$  volts at a constant rate. Thus, the average PD during the charge is  $\frac{V}{2}$  volts. The average power is the product of average PD and the current, i.e.  $P = \frac{V}{2} \times I$  watts. The energy used to charge the capacitor is stored in the charged capacitor and is given by:

$$\text{Energy} = P \times t = \frac{V}{2} \times I \times t$$

$$= \frac{V}{2} \times Q$$

$$= \frac{1}{2} CV^2 \text{ joules}$$

**Capacitive DC Series Circuits**

24. When capacitors are connected in series (Fig 8), the inverse individual values are summed and the reciprocal found to determine the total capacitance of the circuit, i.e.:

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \frac{1}{C_4}$$

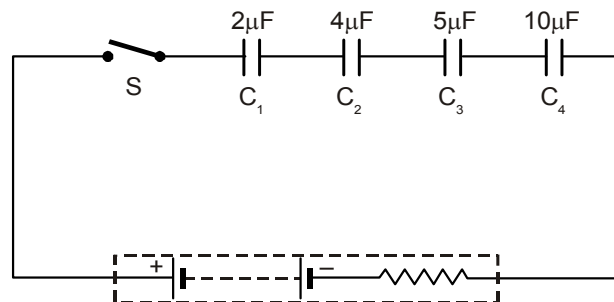
Substituting the values from Fig 8,

$$\frac{1}{C} = \frac{1}{2} + \frac{1}{4} + \frac{1}{5} + \frac{1}{10} = \frac{10}{20} + \frac{5}{20} + \frac{4}{20} + \frac{2}{20} = \frac{21}{20}$$

$$\therefore C \approx 1 \mu\text{F}$$

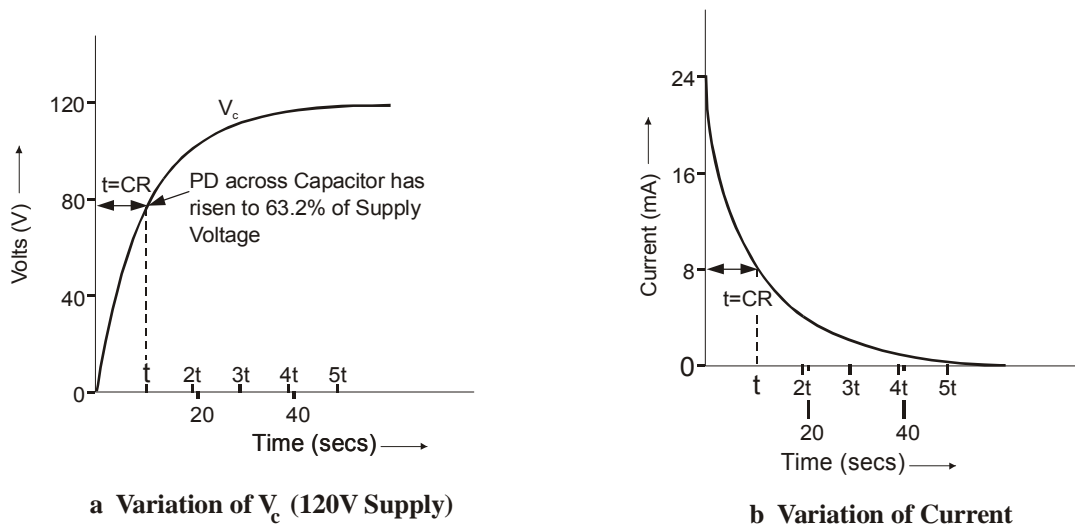
The effect of the series connections is to create a 'total capacitor' of large effective distance (d) between its plates. In Fig 8, the resistor represents the circuit resistance.

**14-4 Fig 8 Capacitors in Series**



25. When only the resistance of connectors is present, the time taken to charge the capacitors is negligible, but becomes significant if a large resistor is connected in series. This is discussed further in para 27. Fig 9a shows a plot of the PD ( $V_c$ ) across the capacitor against time; the equivalent current curve shown at Fig 9b is inverted with respect to the voltage curve.

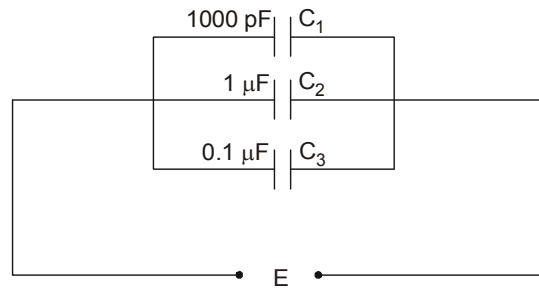
**14-4 Fig 9 Voltage and Current during Charges in a Capacitive DC Circuit**



### Capacitive DC Parallel Circuits

26. The total capacitance of several capacitors in parallel (Fig 10) is the sum of their individual values, ie  $C = C_1 + C_2 + \dots + C_n$ . Substituting the values from Fig 10:

$$C = 0.001 + 1 + 0.1 = 1.101 \mu\text{F}$$

**14-4 Fig 10 Capacitors in Parallel**

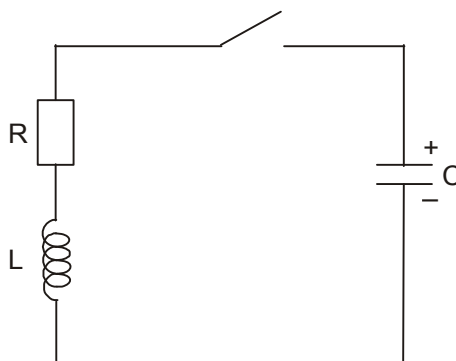
The capacitance is increased because the effective plate area is increased.

### Capacitive-Resistive Circuits, Time Constants

27. The time required to charge a capacitor depends upon both its capacitance and the circuit resistance. The time ( $t$ ) is equal to  $CR$  seconds (where  $C$  is in farads and  $R$  in ohms) and is known as the 'time constant' of a capacitive-resistive circuit. The time constant is defined as the time taken for the PD across a capacitor to rise to 63.2% of the maximum value of the supply voltage on charge, or to fall by 63.2% of its fully charged PD when on discharge. The time constant ( $t$ ) is indicated in Figs 9a and 9b. In theory, a capacitor would take an infinitely long time to completely charge or discharge, but after a time of  $5t$  seconds (ie  $5CR$ ) the charge or discharge is complete for all practical purposes.

### The Oscillatory Circuit

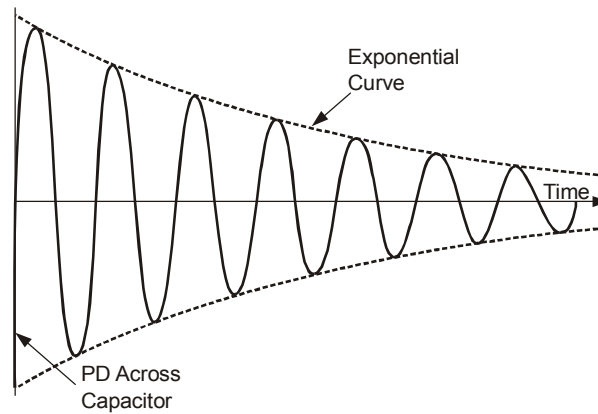
28. In the electrical circuit shown in Fig 11, a charged capacitor is connected in series with an inductor of inductance  $L$  and resistance  $R$ . Closing the switch causes the capacitor to discharge and current flows in the circuit. This current flow generates a magnetic field within the coils of the inductor, and the electrostatic energy stored in the capacitor is transferred to the inductor in the form of electromagnetic energy.

**14-4 Fig 11 The Oscillatory Circuit**

29. Once the capacitor is completely discharged, the magnetic field within the inductor starts to collapse and the collapsing (changing) flux tends to keep the current flowing. The capacitor finally charges up once again, but this time with the reverse polarity to the original charge. When the current has fallen to zero, the capacitor begins to discharge again and the same sequence of events is repeated. This exchange of energy between two parts of the circuit is called 'oscillation', and the period between two successive instants at which the conditions in the circuit are similar is called a 'cycle'.

30. Since the circuit contains resistance, a small amount of energy is dissipated in the form of heat during each cycle. Therefore, each successive charge across the capacitor is smaller in magnitude than the previous one. This effect is known as 'damping' and the decay in amplitude due to damping follows an exponential curve. This is depicted graphically in Fig 12.

**14-4 Fig 12 Oscillatory Damping due to Circuit Resistance**



31. For any given oscillatory circuit, there is a value of resistance which will prevent oscillations taking place. This resistance, known as 'critical damping resistance', is given by the following equation:

$$\text{Critical Damping (R)} = 2\sqrt{\frac{L}{C}}$$

If R equals or exceeds this value, then the nature of the discharge of the capacitor is that of a simple CR combination referred to in para 26. If R is much lower than the critical damping value, then the frequency of oscillation is approximately:

$$\text{Resonant Frequency (f}_r\text{)} = \frac{1}{2\pi\sqrt{LC}}$$

## CHAPTER 5 - MEASURING INSTRUMENTS

Contents	Page
Introduction .....	1
CURRENT AND VOLTAGE .....	2
Ammeters and Voltmeters .....	2
Moving Coil Instruments .....	2
Moving Iron Instruments .....	3
Shunts and Multipliers .....	4
Rectification for AC Measurements .....	4
RESISTANCE .....	5
Volt-amp Method .....	5
The Ohmmeter .....	5
The Wheatstone Bridge Method .....	6
POWER .....	6
Volt-amp Method .....	6
Wattmeters .....	7
OTHER COMMON INSTRUMENTS .....	7
The Electrostatic Voltmeter .....	7
The Thermo-ammeter .....	8
Digital Measuring Instruments .....	8
Flux Meters .....	8

### Table of Figures

14-5 Fig 1 Ammeter and Voltmeter Connections in a Circuit .....	2
14-5 Fig 2 A Simple Moving Coil Instrument .....	3
14-5 Fig 3 A Moving Iron Instrument .....	4
14-5 Fig 4 Rectifier and Moving Coil Instrument .....	5
14-5 Fig 5 A Simple Ohmmeter .....	5
14-5 Fig 6 A Wheatstone Bridge .....	6
14-5 Fig 7 A Wattmeter .....	7
14-5 Fig 8 The Electrostatic Voltmeter .....	7
14-5 Fig 9 The Thermo-ammeter (Principle of Operation) .....	8
14-5 Fig 10 Fluxgate Core Windings .....	9

### Introduction

1. In all fields of science and engineering instruments are required in order to measure the quantities being used; electrical engineering is no exception to this. Associated with most electrical circuits are the basic quantities of voltage, current, power and resistance. The instruments which measure these are the voltmeter, ammeter, wattmeter, and ohmmeter respectively.



## CURRENT AND VOLTAGE

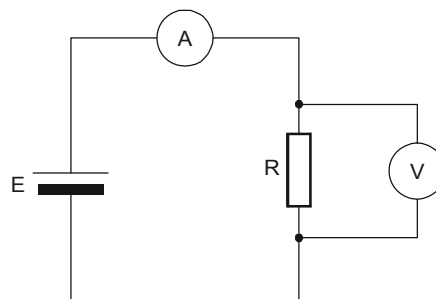
### Ammeters and Voltmeters

2. An ammeter is used to measure current flow through a circuit and therefore the instrument must be connected in series with the circuit, as shown in Fig 1. It is important that the current being measured is unaffected by the instrument's internal resistance, and for this reason ammeters are designed to have a low electrical resistance. This being the case, an ammeter should never be connected across a supply voltage, since there would be nothing to limit the current flow and the instrument would be damaged.

3. A voltmeter is designed to measure electromotive force and potential difference. The potential difference across a resistor is measured by connecting the meter as shown in Fig 1. In order to ensure the minimum disturbance in any circuit voltmeters are required to pass very little current and therefore have a high internal resistance.

4. There are a number of ways that an electric current or voltage can cause the needle of a measuring instrument to be deflected. The most common method utilizes the magnetic field of an electric current to produce the deflection. In this case it is important to realize that regardless of the units being measured the principle of operation remains the same; ie current through the instrument causes needle deflection. The internal resistance, high or low relative to the circuit under test, determines whether the device is a voltmeter or an ammeter.

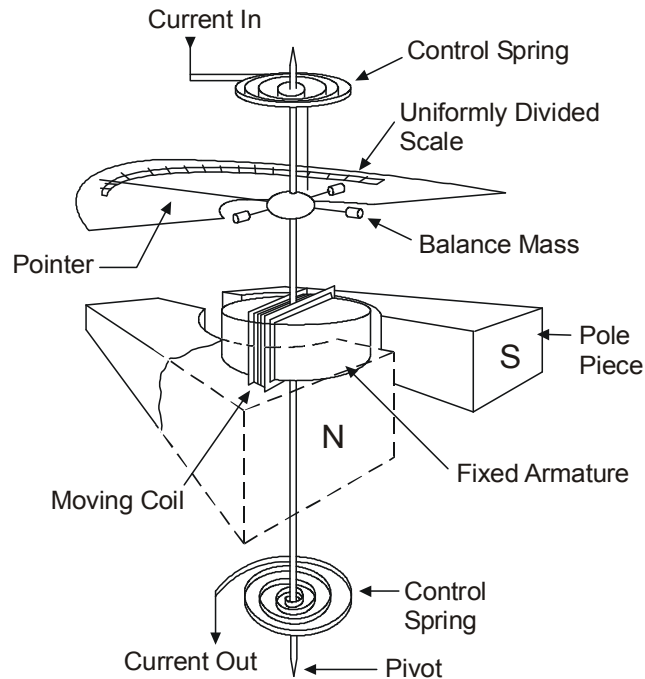
**14-5 Fig 1 Ammeter and Voltmeter Connections in a Circuit**



### Moving Coil Instruments

5. The operation of a moving coil instrument depends on the reaction between the current in a moveable coil and the field of a fixed permanent magnet. This reaction is discussed in more detail in Volume 14, Chapter 6, under the topic of DC motors.

6. The moving coil is wound around an aluminium former and suspended between the curved poles of a magnet, as shown in Fig 2. The current to be measured is fed through the moveable coil via two coiled springs. A magnetic field is set up around the moveable coil which reacts with that of the magnet and produces a turning force (torque). The coil will turn and apply tension to the coiled springs until the restoring torque of the springs is equal to the force produced by the coil. The amount of angular movement is proportional to the applied current and is indicated by a pointer on a scale.

**14-5 Fig 2 A Simple Moving Coil Instrument**

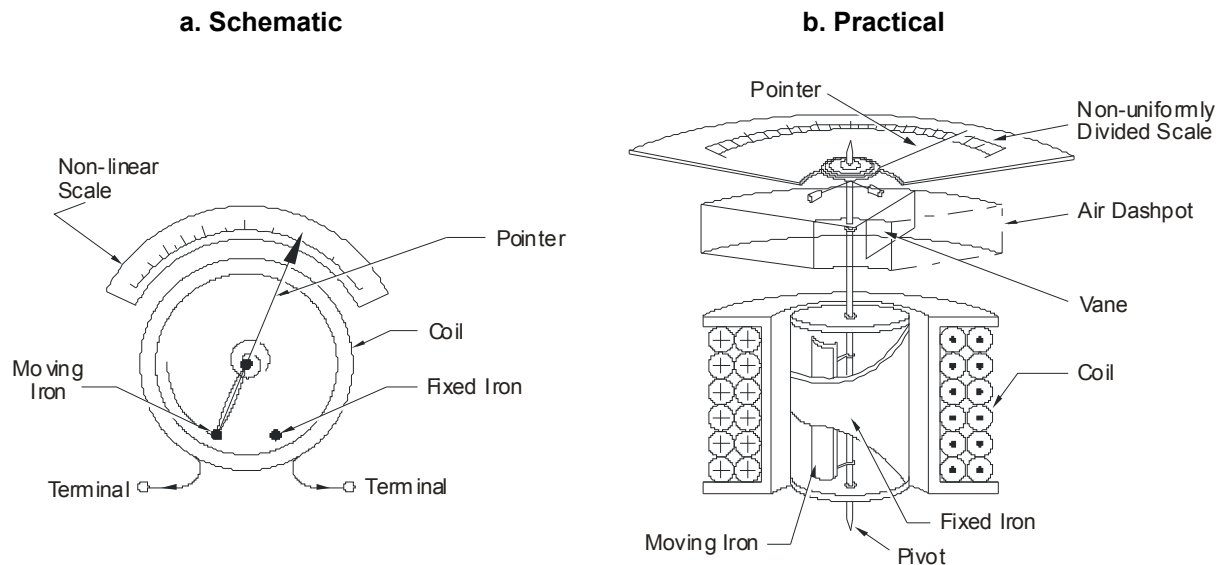
7. In order to prevent the pointer overshooting and then oscillating before settling down at the correct reading, damping of the movement must be provided. This is achieved through the aluminium former supporting the coil. When the coil moves through the magnetic field, small circulating currents (eddy currents) are set up in the former which produce a restraining torque while the coil is in motion and thus damp out oscillation.

8. The main features of a moving coil instrument are:

- a. It measures DC only.
- b. A linear scale.
- c. High sensitivity.
- d. It may be used as a DC ammeter with a shunt (para 11).
- e. It may be used as a DC voltmeter with a multiplier (para 12).

### **Moving Iron Instruments**

9. The repulsion moving iron instrument is basically two pieces of iron placed inside a coil, as shown in Fig 3a. One of the pieces is fixed while the other is attached to a pointer and free to move. When current is passed through the coil, a magnetic field is produced which magnetizes both irons in the same sense. Since like poles repel each other, the free iron is forced away from the fixed iron. As in the case of the moving coil instrument, a coiled spring at the pivot point of the pointer provides the restoring torque.

**14-5 Fig 3 A Moving Iron Instrument**

10. In the practical instrument, Fig 3b, the moving iron is rectangular in shape, while the fixed iron is in the shape of a tapered scroll. This is done to improve the scale linearity due to the non-linear forces between two irons of similar shape. Instrument damping is achieved by means of a vane and air dashpot. The main features of this type of instrument are:

- a. It measures DC and AC.
- b. A non-linear scale.
- c. It may be used as an ammeter, but shunts should not be used, due to their non-linearity.
- d. It may be used as a voltmeter, directly or with multipliers.
- e. It is accurate only for frequencies below about 200 Hz.

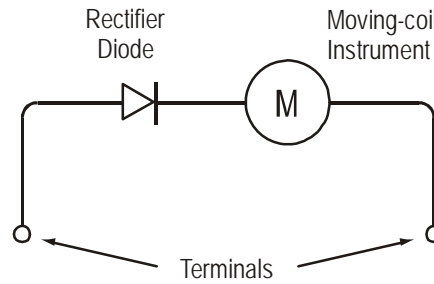
**Shunts and Multipliers**

11. In general, meters are constructed to have a high sensitivity and therefore give full-scale deflection (fsd) of the pointer with small currents of the order of a few micro-amps. In order to measure currents of a higher magnitude, resistors, known as shunts, are connected across the meter in order to by-pass the excess current. The meter may now be calibrated to read an apparent fsd which is equal to the inherent fsd plus the excess current through the resistor.

12. If a meter is required to read voltages then a resistor has to be connected in series with the instrument across which most of the applied voltage is developed. The voltage across the instrument itself is small and the current flowing through it is still only of the order of micro-amps. This type of resistor is referred to as a multiplier.

**Rectification for AC Measurements**

13. A rectifier is a device which allows current to flow in one direction only, very much like a non-return valve in a fluid system. When such a device is connected in series with a moving coil instrument, as shown in Fig 4, a rectifier allows the instrument to be used for the measurement of AC voltages and currents.

**14-5 Fig 4 Rectifier and Moving Coil Instrument**

14. The instrument indicates the average value of the current flowing through it, and thus gives a reading proportional to the average value of the half-wave rectified waveform. Meters of this type give accurate readings for AC voltages and currents up to about 20 kHz, but at low frequencies the pointer tends to fluctuate.

**RESISTANCE****Volt-amp Method**

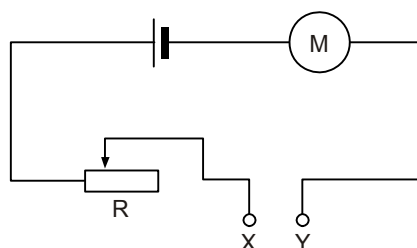
15. The value of a resistor (R) may be determined by voltage (V) and current (I) measurements using instruments arranged as shown in Fig 1. By applying Ohm's law to the values obtained the resistance may be calculated.

$$R = \frac{V}{I} \text{ ohms}$$

**The Ohmmeter**

16. The disadvantage of the volt-amp method of determining resistance is the need for two meters, a voltmeter and an ammeter. This problem is overcome with the introduction of a single meter known as the ohmmeter. Basically, the ohmmeter uses the same movement as the voltmeter and the ammeter, but the zero reading is at the opposite end of the scale.

17. The circuit of a basic ohmmeter is shown in Fig 5. It comprises a moving coil ammeter (M), a variable resistor (R) and a battery. Prior to any measurement being taken the terminals X and Y are shorted together and the variable resistor adjusted to give full-scale deflection of the needle; this corresponds to zero resistance. Any resistance placed between the two terminals will cause less current to flow in the circuit and result in less deflection of the needle. The amount of deflection is therefore inversely proportional to the value of the resistance.

**14-5 Fig 5 A Simple Ohmmeter**

### The Wheatstone Bridge Method

18. Bridge methods of measuring component values are used where high accuracy is most important. The most straightforward bridge circuit is the Wheatstone bridge. It is less convenient than an ohmmeter, but far more accurate.

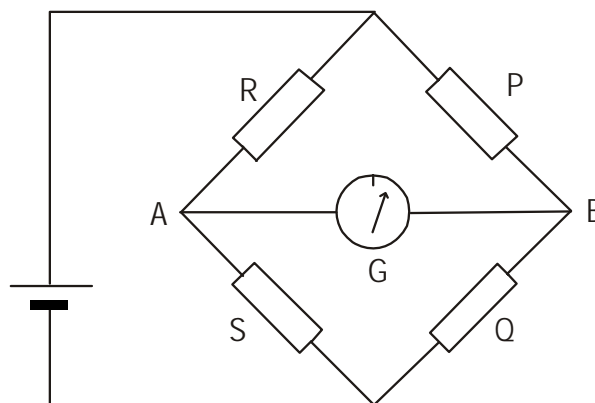
19. The basic configuration consists of a four limb resistance bridge, as shown in Fig 6, where R is the resistance to be measured, and resistors P, Q and S are variable standard resistors.

20. The variable resistors (P, Q and S) are adjusted until no current is registered in the galvanometer connected between points A and B (a galvanometer is a very sensitive DC current measuring instrument using the same principle as the moving coil instrument). Under these conditions the bridge is said to be in balance and the ratio of P to Q will then equal that of R to S.

Therefore, 
$$\frac{P}{Q} = \frac{R}{S}$$

$$R = \frac{P}{Q} \times S$$

**14-5 Fig 6 A Wheatstone Bridge**



### POWER

#### Volt-amp Method

21. Referring back to the circuit shown in Fig 1, power in the resistor may be found by measuring the voltage across the resistor and the current through it. The power (P) is then given by the equation:

$$P = VI$$

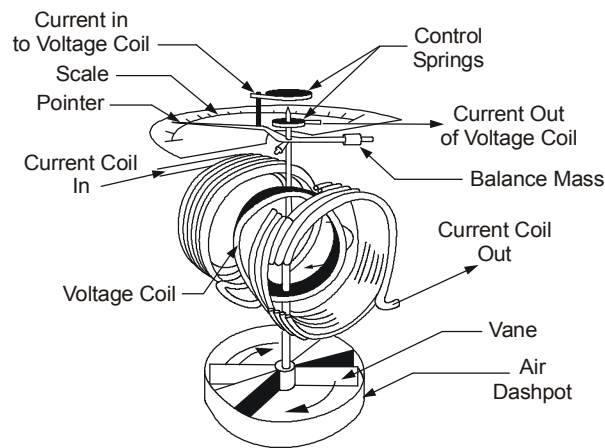
22. This method may be used to give accurate measurement of power in a resistive load in AC and DC circuits. Where loads are not purely resistive, ie they contain inductance and capacitance, the foregoing method of determining power gives incorrect results; this is due to the phase differences between voltage and current. Under these conditions a wattmeter is used.

## Wattmeters

23. The moving coil meters discussed in the earlier paragraphs of this chapter rely on permanent magnets to provide the fixed magnetic field. However, this magnetic field can be created by passing current through fixed current coils. These additional coils are the basis of wattmeter design.

24. A wattmeter consists of two fixed low resistance current coils and a high resistance voltage coil which is free to move as shown in Fig 7. The amount of pointer deflection will be proportional to the current flowing through the load and the voltage across it; ie power.

**14-5 Fig 7 A Wattmeter**

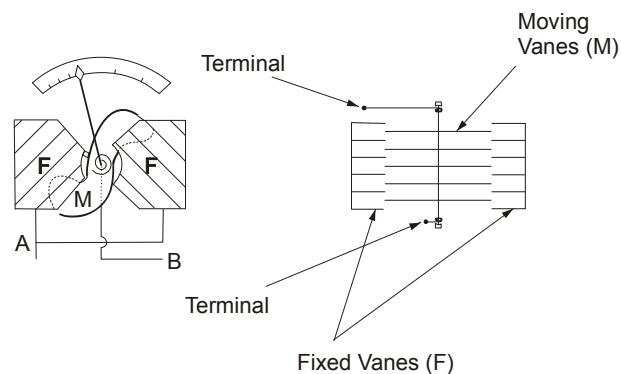


## OTHER COMMON INSTRUMENTS

### The Electrostatic Voltmeter

25. The electrostatic voltmeter is essentially a variable capacitor whose moving plates are suspended by a spring or wire and are free to move, as shown in Fig 8. When a voltage is applied, the mutual attraction between the fixed and moving plates causes the latter to turn until restrained by the restoring force of the spring. Since no current flows after the initial charging of the plates, the electrostatic voltmeter consumes no power and has a very high internal resistance-ideal for voltage measurement. The chief disadvantage of this type of meter is its insensitivity over the low voltage ranges.

**14-5 Fig 8 The Electrostatic Voltmeter**

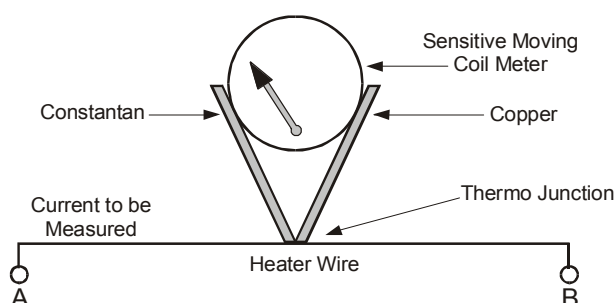


## The Thermo-ammeter

26. When two dissimilar metals are in contact with each other an electromotive force (emf) is produced at the point of contact. The magnitude of this emf depends on the metals used and the temperature of the junction. This principle is used in the thermo-ammeter, an instrument which can be used for AC measurement at both low and high frequencies.

27. In the simplified diagram, shown in Fig 9, current passing through the wire heats the junction of the thermo-couple. The resulting emf can then be measured using a standard moving coil instrument. In addition to being a most useful instrument for high frequency current measurements, the thermo-ammeter principle is used extensively in engine temperature monitoring.

**14-5 Fig 9 The Thermo-ammeter (Principle of Operation)**



## Digital Measuring Instruments

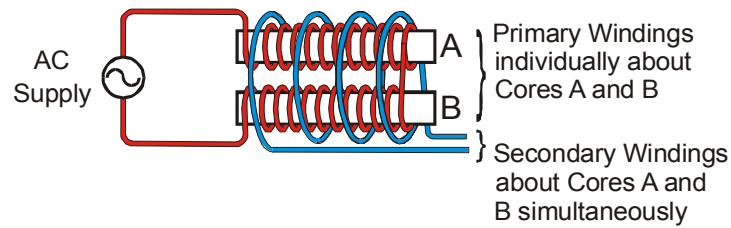
28. The instruments discussed so far indicate their readings by the movement of a pointer over a scale and are known as 'analogue' instruments. An alternative form of display presents the reading as a sequence of digits, and instruments with this form of display are referred to as 'digital' instruments. Digital measuring instruments have the advantage that the measured quantity is displayed in numerical form and, in comparison to the moving pointer type of instrument, they are quicker and easier to read. When used to measure voltages, digital instruments are usually referred to as DVMs (digital voltmeters).

29. The reading is produced by a voltage measuring analogue-to-digital converter. The binary output from the circuit is applied to a decoder which drives the meter display. When the input voltage is of the order of millivolts, it is normally amplified before being measured. For varying voltages, a latch circuit is included to hold the display steady at the last value while further samples of the input are taken. With some extra internal components, currents and resistances can be measured, thus allowing the various multimeter tasks to be undertaken.

## Flux Meters

30. The need to measure the strength of magnetic fields in science and industry has resulted in the production of magnetometers of many different types. Early magnetometers consisted of spring-loaded magnets in which field strength was measured by the extension of the spring. These were soon superseded by electronic systems, one of the more common being the fluxgate magnetometer.

31. The fluxgate magnetometer depends for its operation on the rapid AC magnetization of a pair of saturable cores. Exciter coils are wound round each of the cores and are in phase opposition. The output sum is taken via a secondary winding which is wound round the outside of both cores. This is depicted in Fig 10.

**14-5 Fig 10 Fluxgate Core Windings**

32. Under external flux free conditions, the magnetic flux induced by the AC excitation in one core cancels out the flux induced in the other; the resultant being a zero output. The introduction of an external magnetic field has the effect of introducing magnetic bias into the system and causes one of the cores, depending on sense, to saturate. Core saturation results in the flux in one core having a greater influence than that in the other. The summing effect of the output coil produces AC current, of twice the input frequency, which is proportional in magnitude to the external magnetic field. This may be measured using a moving coil meter and rectification.

33. Fluxgate magnetometers are capable of detecting rapidly varying magnetic fields, and are ideally suited to the detection of submarines, and sunken objects such as ships or mines.



## CHAPTER 6 - DC MACHINES

Contents	Page
Introduction .....	2
The Simple Generator .....	2
DC GENERATORS .....	4
Fleming's Right-hand Rule .....	4
Production of DC .....	5
Improving the DC Output.....	6
Variation of the Magnitude of the Induced EMF .....	7
Practical DC Generator .....	7
Classification of DC Generators .....	8
Generator Losses and Efficiencies .....	10
THE DC MOTOR.....	11
General Principles .....	11
Fleming's Left-hand Rule .....	12
The Simple Motor .....	12
The Practical Motor .....	13
The Speed of a Motor.....	13
Types of Motors.....	15
Motor Starters.....	16
Rotary Transformers .....	16

### Table of Figures

14-6 Fig 1 The Simple Generator.....	2
14-6 Fig 2 Values of Induced Voltage for One Revolution of the Coil.....	3
14-6 Fig 3 Fleming's Right-hand Rule.....	4
14-6 Fig 4 Production of DC by Commutator Action.....	5
14-6 Fig 5 A Two-loop Generator.....	6
14-6 Fig 6 Two-loop Generator Output .....	6
14-6 Fig 7 A Typical DC Generator .....	7
14-6 Fig 8 Shunt Generator.....	9
14-6 Fig 9 Series Generator .....	9
14-6 Fig 10 Production of Eddy Currents .....	10
14-6 Fig 11 Splitting Eddy Currents by Laminations .....	10
14-6 Fig 12 Losses and Efficiency of a Compound Generato .....	11
14-6 Fig 13 Force on a Current Carrying Conductor .....	12
14-6 Fig 14 Fleming's Left-hand Rule .....	12
14-6 Fig 15 DC Motor Principle .....	13
14-6 Fig 16 Speed Varied by Controlling Field Strength.....	14
14-6 Fig 17 Speed Varied by Controlling Armature Current .....	14
14-6 Fig 18 A Shunt-wound Motor .....	15
14-6 Fig 19 A Series-wound Motor.....	16

## Introduction

1. While a conductor is moving in a magnetic field in such a way that the magnetic flux linkage is changing, there is in that conductor an induced emf which at any instant is proportional to the rate of change of flux linkage (Volume 14, Chapter 3). If this emf is applied to a closed circuit an electric current will be established. Any machine which produces such an emf is known as a generator. The DC generator (or dynamo) is a mechanically driven machine in which conductors rotating in a magnetic field generate a DC output voltage.

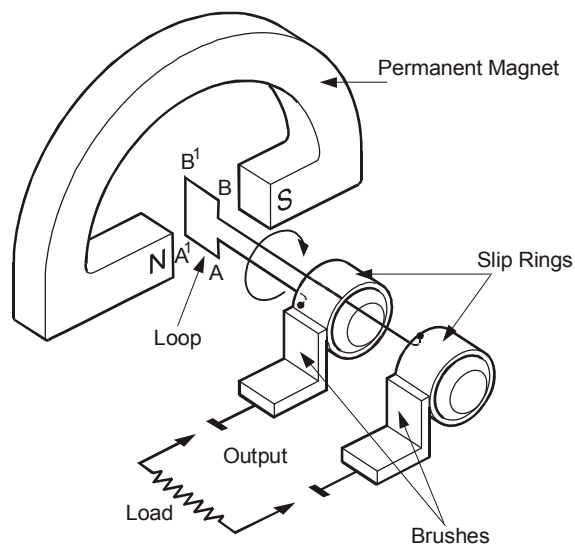
2. An electric motor converts electrical energy into mechanical energy, its action being the reverse of that of the generator. There is little difference in the design of generators and motors; providing the connections are suitable a DC machine can be used as either a generator or a motor.

## The Simple Generator

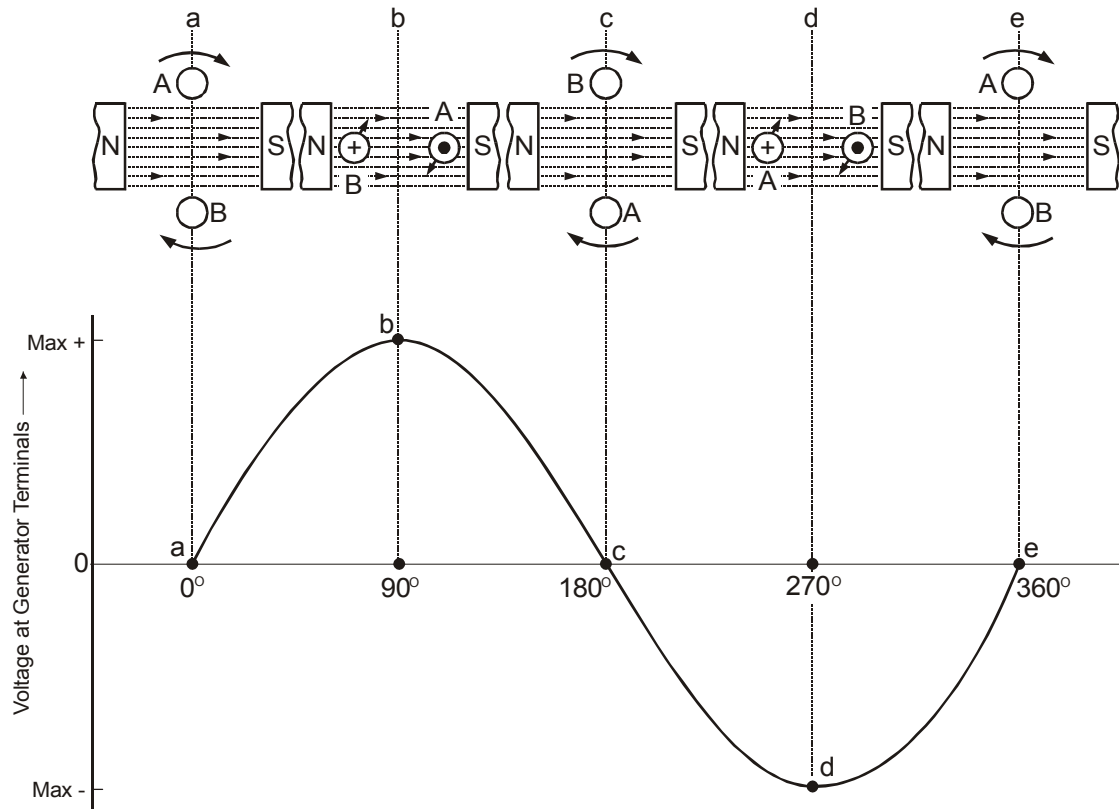
3. The simplest form of generator consists of a single loop of wire which can be rotated freely in the space between the poles of a permanent magnet. Connection is made to the external circuit (or 'load') by brushes pressing on two slip rings connected to the ends of the loop (Fig 1).

4. While the loop is rotating the magnetic flux linkage is changing, and an emf will be induced in each of the straight sides  $A - A^1$  and  $B - B^1$ . The direction of the induced emfs given by Fleming's Right-hand Rule (see para 8) is such that the two emfs in series are additive and combine to establish a current when the load is connected.

14-6 Fig 1 The Simple Generator



5. The magnitude of the induced emf is proportional to the rate of change of flux linkage; thus, assuming the speed of rotation to be constant, the induced emf will at any instant depend on the position of the loop in the magnetic field. This is illustrated in Fig 2, which represents the view from the slip ring end of the loop.

**14-6 Fig 2 Values of Induced Voltage for One Revolution of the Coil**

a. In position 'a' Fig 2, the conductors A and B are moving parallel to the lines of magnetic flux. In a very small period of time  $dt$  about the instant shown in 'a', the change of flux linkage ( $d\Phi$ ) is zero. Thus the rate of change of flux linkage  $\frac{d\Phi}{dt}$  is zero, and since  $E$  (the back emf) =  $-\frac{d\Phi}{dt}$  volts, the emf induced in the loop at this instant is zero.

b. Position 'b' shows the conductors cutting the flux at right angles. The rate at which the flux linkage is changing is now a maximum and the emf induced in the loop is a maximum, the direction being given by Fleming's Right-hand Rule. At this position the emf is arbitrarily assumed to be maximum in a positive direction.

c. Position 'c' represents a position where the rate of change of flux linkage is again zero and no emf is induced in the loop.

d. Position 'd' is similar to position 'b' except that the side of the loop which was previously moving downwards (side A) is now moving upwards and vice versa. The rate of change of flux linkage is again a maximum and Fleming's Right-hand Rule will confirm that in relation to 'b' the emf is now a maximum in a negative direction.

e. Position 'e' is identical with position 'a'.

6. At other positions of the loop, the rate of change of flux linkage is intermediate between zero and maximum and so, therefore, is the emf. Thus during one complete revolution of the loop the voltage at the terminals of the generator will vary in the manner shown in the graph of Fig 2. This shows one cycle of alternating voltage.

7. The frequency of the alternating voltage in cycles per second, or hertz, at the generator terminals depends on the speed of rotation and on the number of pairs of poles in the field magnet system. Thus:

$$\text{frequency } f = \frac{pN}{60} \text{ (Hz)}$$

where  $p$  = number of pairs of poles and  $N$  = revolutions per minute

Hence the frequency of an alternating voltage produced by a 4-pole AC generator running at 1,500 rpm is:

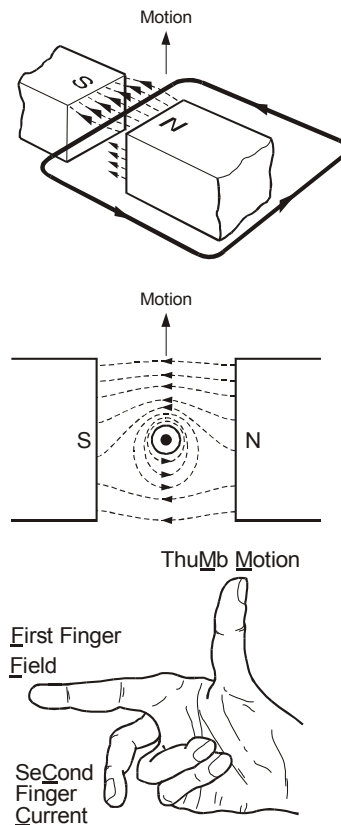
$$f = \frac{pN}{60} = \frac{2 \times 1,500}{60} = 50 \text{ Hz}$$

## DC GENERATORS

### Fleming's Right-hand Rule

8. The right-hand rule for generators is a method of remembering the direction of the voltage induced in a conductor moving in a magnetic field. The 'direction of voltage' indicates the direction in which conventional current would flow in a closed circuit as a result of this voltage. The right-hand rule (Fig 3), sometimes called 'the geneRIGHTer rule', requires the thumb, the first finger and the second finger of the RIGHT hand to be held at right angles to each other. With the thumb pointing in the direction in which the conductor has been Moved, and the First finger in the direction of the magnetic Field (N to S), the seCond finger indicates the direction in which conventional Current would flow in the conductor; this in turn gives the direction of the induced voltage.

14-6 Fig 3 Fleming's Right-hand Rule

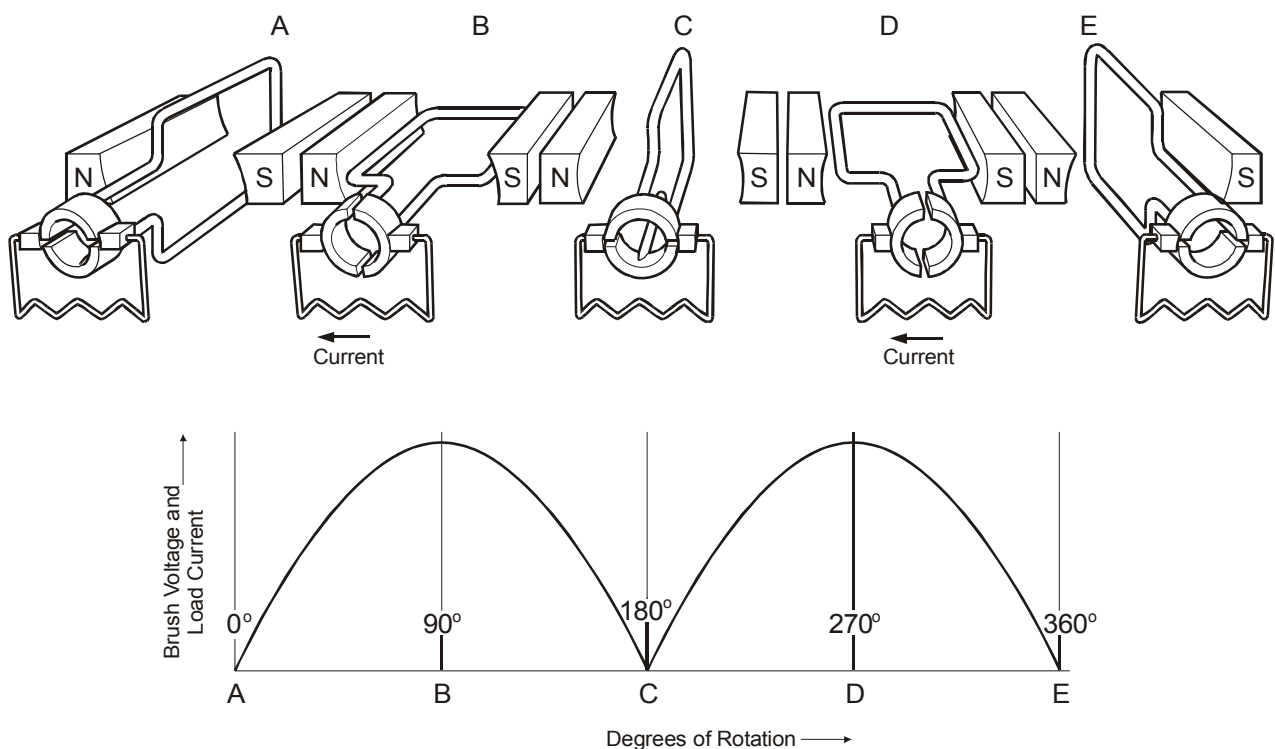


## Production of DC

9. The alternating voltage produced in the loop reverses its polarity every time the loop goes through the  $0^\circ$  and  $180^\circ$  positions, because at these points the conductors forming the loop reverse their direction of travel across the magnetic field. To obtain DC from the generator, the AC produced by the rotating loop must be converted to DC. A possible method is to have a switch connected across the output in such a way that the connections to the load are reversed every time the polarity of the voltage in the loop changes. The commutator is a device which does this, switching automatically as the loop rotates, thus maintaining the same direction of current in the load.

10. **The Commutator.** A simple commutator for a single-loop generator consists of the two halves of a slip ring insulated from each other and from the shaft which carries them and the loop. Each end of the loop is connected to a segment of the commutator, and the load is connected to the loop by brushes bearing on opposite sides of the commutator as shown in Fig 4. As the loop rotates, an alternating voltage is induced in it; a fluctuating DC output is derived from this alternating voltage by using a commutator. Since the commutator rotates with the loop, the brushes bear on opposite segments during each half cycle (compare B and D of Fig 4). The left-hand brush is always in contact with that segment which is negative, and the right-hand brush with that segment which is positive. The changeover from one segment to the other takes place at the instants when the voltage induced in the loop is zero, at A, C and E. The current in the external circuit is therefore always in the same direction, it is a unidirectional current, and this is the first step towards obtaining a true DC output such as we get from a battery. The variations in brush voltage and external circuit current during one complete revolution of the loop are also shown in Fig 4. Note that within the rotating loop itself an alternating voltage is produced; the commutator is merely a switch that reverses the direction of alternate half-cycles in the output leads.

14-6 Fig 4 Production of DC by Commutator Action

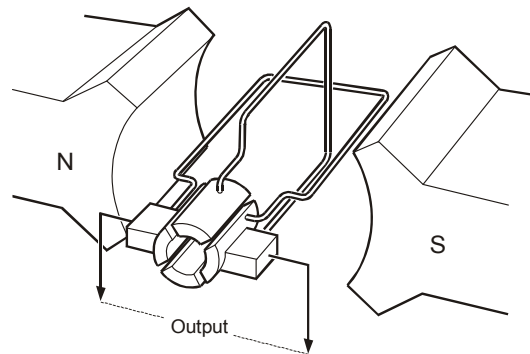


## Improving the DC Output

11. The voltage at the brushes and the current in the external circuit of a single-loop DC generator fall to zero twice during each revolution. The voltage wave-form is improved, so that it becomes smoother and of greater amplitude, by:

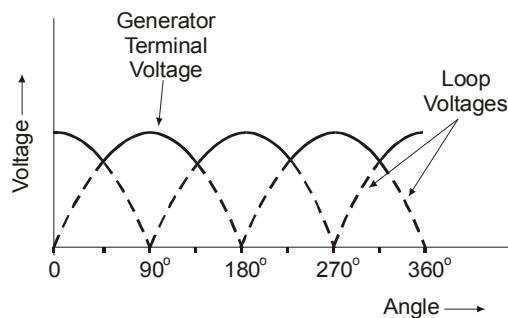
- Increasing the numbers of loops, which necessitates additional segments on the commutator.
- Creating stronger magnetic fields.

**14-6 Fig 5 A Two-loop Generator**



12. **Use of More Than One Loop.** When just one loop is added, thus giving two loops at right angles to each other (as shown in Fig 5), the output voltage using a commutator with four segments becomes that shown in Fig 6. The ripple that remains may be made smoother by the use of additional symmetrically placed loops, and the use of a commutator with a correspondingly large number of segments.

**14-6 Fig 6 Two-loop Generator Output**



13. **The Creation of Stronger Magnetic Fields.** Greater effective strength of the magnetic fields is obtained by:

- Using electromagnets instead of permanent magnets. The pole pieces are made of soft iron and are magnetized by a current flowing through the coils wound round them. These are known as the field windings.
- Winding the rotating loops, which should be made of many turns of wire, on a soft iron core to concentrate the magnetic field through the loops. This assembly is known as the armature. The coils, are, of course, insulated from the iron core.
- Decreasing the air gap between the poles and armature by shaping them to fit each other.
- Increasing the number of poles. Four pole generators are quite usual; heavy duty and larger machines use more.

### Variation of the Magnitude of the Induced EMF

14. From Faraday's and Lenz's laws the emf induced in a conductor moving through a magnetic field is given by:

$$E = -n \frac{d\Phi}{dt} \text{ volts}$$

where  $\frac{d\Phi}{dt}$  represents the rate of change of flux, and  $n$  the number of turns of the coil. The emf of a generator therefore depends on:

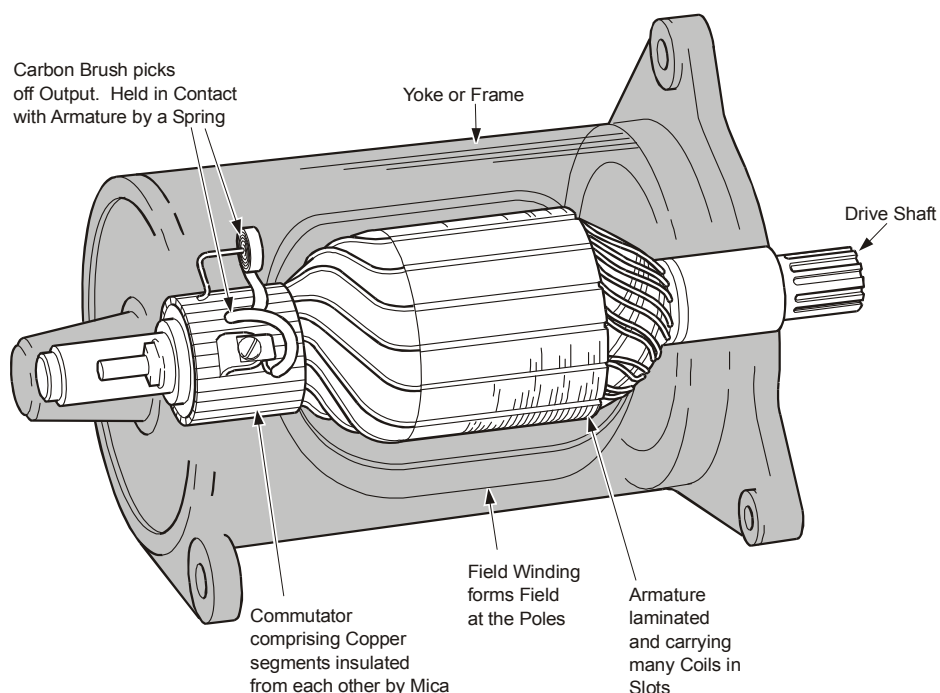
- The number of conductors connected in series between the brushes.
- The speed of rotation.
- The magnetic flux density.

15. In practice the number of conductors is fixed and the speed is nominally constant, so control of the emf must be obtained by variation of the magnetic flux density. Using permanent magnets the flux density is constant, but by substituting electromagnets the emf can be controlled by varying the current in the magnet windings.

### Practical DC Generator

16. A small DC generator is illustrated in Fig 7. The mainframe or yoke is the main chassis of the generator and it also serves to complete the magnetic circuit between the pole pieces. The pole pieces are laminated to reduce eddy current losses (see para 25), and the field windings are mounted on the pole pieces as shown. The end housings contain the bearings for the armature which rotates at high speed.

**14-6 Fig 7 A Typical DC Generator**



17. The armature is made up of a shaft, armature core, armature windings and the commutator. The armature core is laminated to reduce eddy current losses, the armature winding resting in slots cut in the core but insulated from it. The commutator is made up of copper segments insulated from each other, and from the shaft, by mica. The ends of the armature windings are soldered to their appropriate commutator segments.

18. **The Brush Assembly.** The output emf is picked off from the commutator by the brushes. They are made of some form of carbon which is self-lubricating and which causes very little wear of the commutator. In addition their relatively high resistance minimizes a tendency for sparking to occur as the brushes pass from one commutator segment to the next. The carbon pieces are mounted in brush holders and are held against the commutator by tensioned springs. The holders are bolted to brush-rockers which can be adjusted to move a few degrees in either direction so that the tendency for sparking to occur may be further reduced. The brushes are connected to the external circuit by copper wires.

### Classification of DC Generators

19. DC generators are usually classified according to the method by which the magnetic circuit of the machine is energized. The recognized classes are:

- a. Permanent magnet generators.
- b. Separately-excited generators.
- c. Self-excited generators.

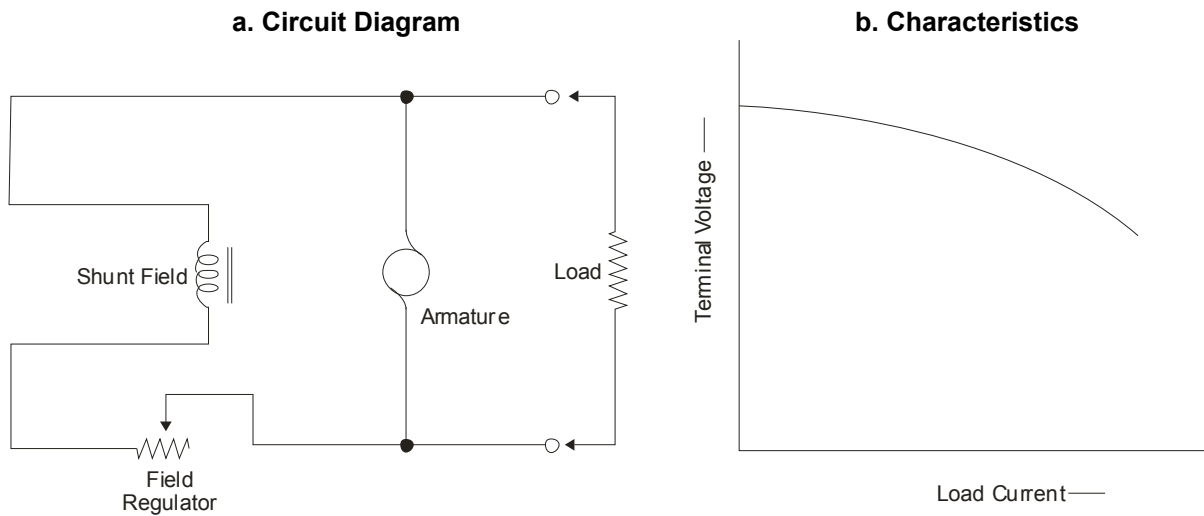
20. **Permanent Magnet Generators.** The permanent magnet type of generator is usually a simple, small machine because of the low strength of the magnetic field. The PD across the terminals decreases slightly as the loading increases, the generator having what is termed a falling external voltage characteristic.

21. **Separately-excited Generators.** The magnetic field for the separately-excited type of generator is obtained from electromagnets which are excited by an external DC source. The exciter supply can be regulated, varying the magnetic field strength, and thus the terminal PD. The terminal PD falls with an increase in loading.

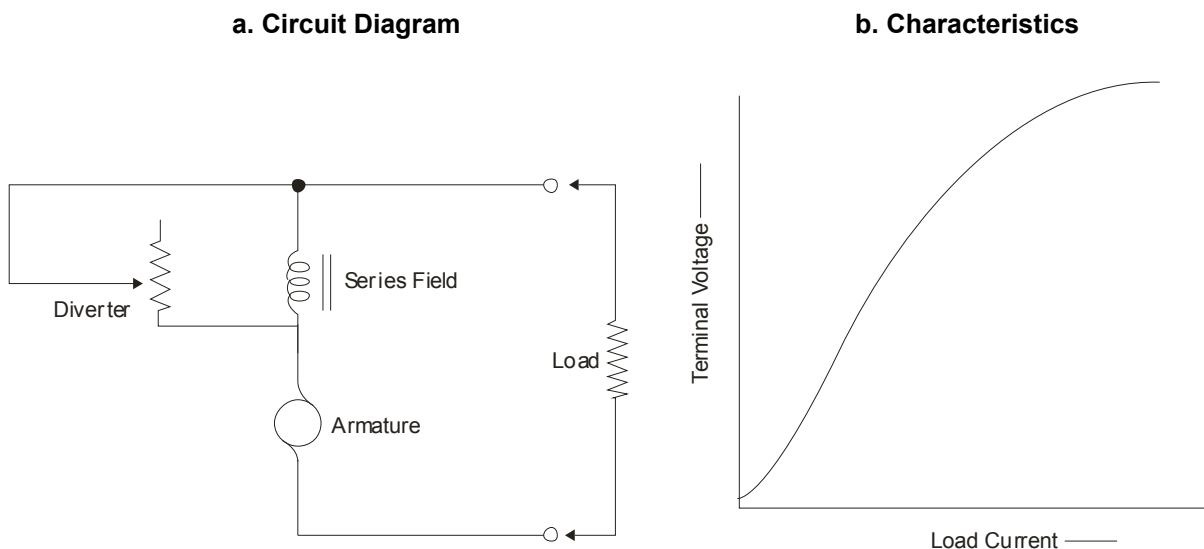
22. **Self-excited Generators.** The power required to excite the electromagnets in the self-excited type of generator is generated by the machine itself. This type of generator may be further classified as follows:

- a. The shunt-wound type, where the field winding is connected directly across the armature. Provision is made for regulating the exciting current by a variable resistor connected in series with the field winding as shown in Fig 8a. Thus control over the terminal PD is obtained. In a shunt generator, as the load draws more current from the armature, the terminal voltage decreases because there is a greater potential drop in the armature and hence a lower terminal PD, and thus field current decreases. The characteristic curve for a shunt generator thus starts at a maximum and then falls as shown in Fig 8b. Note that the generator is excited even with the load disconnected.



**14-6 Fig 8 Shunt Generator**

b. The series-wound type where the field winding is connected in series with the armature and the load. The terminal PD may be controlled by a diverter (a variable resistor connected in parallel with the field winding so as to adjust the current in the field coils). The circuit is shown at Fig 9a. If the load is disconnected no current flows through the field coils; the only voltage induced is that due to residual magnetism. With a load connected, current flows in the armature, the field strength increases and the terminal voltage rises. This continues as the load draws more current until the magnetic field reaches saturation. The characteristic curve for a series generator is shown at Fig 9b.

**14-6 Fig 9 Series Generator**

c. The compound-wound type, which is a combination of the types in sub-paras a and b, and may be constructed so as to give an almost constant voltage output under all loads.

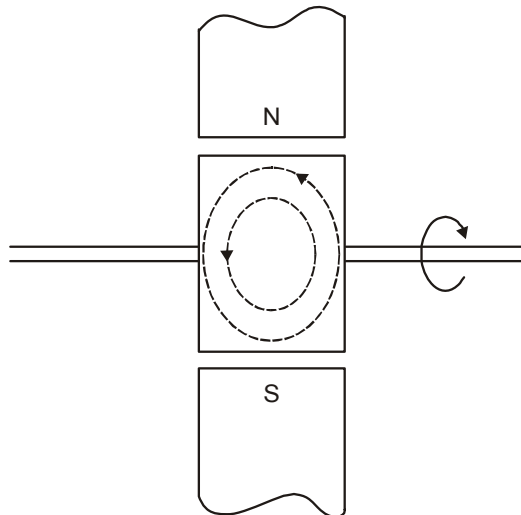
### Generator Losses and Efficiencies

23. There are four types of losses associated with DC generators: copper losses (para 24), hysteresis loss (see Volume 14, Chapter 2, para 23), eddy current loss (para 25), and friction losses (which are self-explanatory).

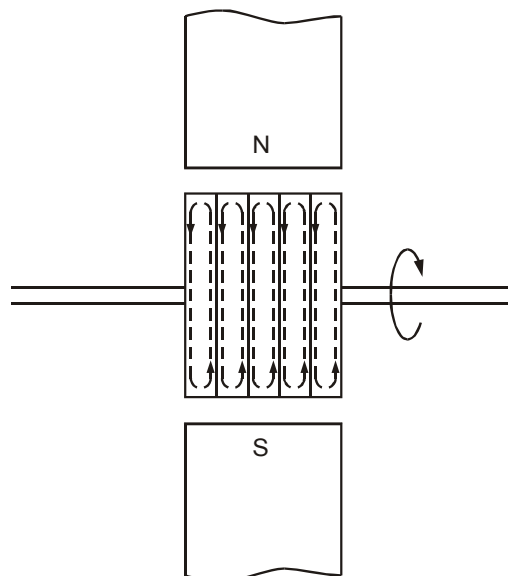
24. **Copper Losses.** Another name for copper losses is electrical losses. They are a power loss ( $I^2R$  watts), caused by the resistance of the conductors. Usually they increase with loading.

25. **Eddy Current Losses.** If the armature core were made of solid iron, as shown in Fig 10 for a two-pole machine, and rotated, eddy currents would be set up in the core. The eddy currents are caused by emfs which are generated in the iron in exactly the same way as emfs generated in conductors wound on the iron core (Fleming's Right-hand Rule will give the current direction). These currents cause two losses: one through the heat generated, and the other by creating magnetic fields which oppose the pole magnetic fields. The losses are minimised by splitting up the solid iron core, ie by laminating it, so that the eddy currents are split also (see Fig 11). A five-lamination core suffers about 1/25th the loss of a solid core ( $\text{loss} \propto I^2$ ).

14-6 Fig 10 Production of Eddy Currents



14-6 Fig 11 Splitting Eddy Currents by Laminations



26. **Efficiencies.** Due to the losses mentioned in the previous paragraph the efficiency of a DC generator in converting mechanical energy into electrical energy varies with the type of machine. Three definitions of efficiency follow:

a. **Mechanical Efficiency.** The electrical power in the armature is equal to the applied mechanical power less the hysteresis, eddy current and friction losses. Thus:

$$\text{mechanical efficiency (\%)} = \frac{\text{power in armature}}{\text{power supplied}} \times 100$$

b. **Electrical Efficiency.** The electrical power output is equal to the power in the armature less the copper losses. Thus:

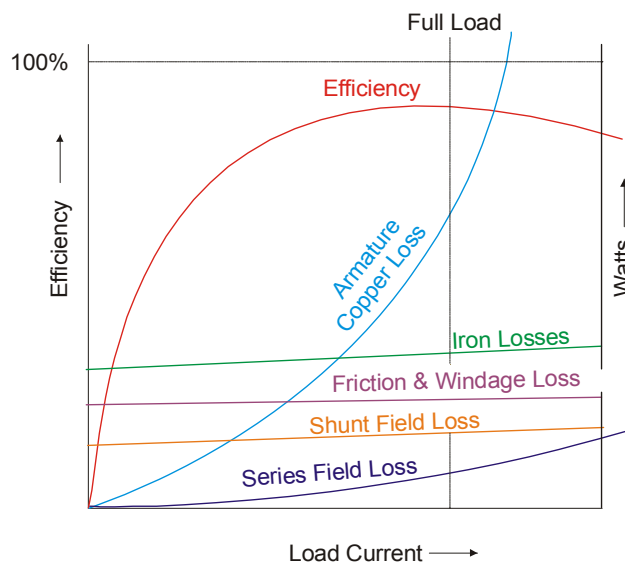
$$\text{electrical power output efficiency (\%)} = \frac{\text{power output}}{\text{power in armature}} \times 100$$

c. **Commercial Efficiency.** The commercial efficiency is the one most commonly used to describe the overall efficiency of a machine and:

$$\text{commercial efficiency (\%)} = \frac{\text{electrical power output}}{\text{mechanical power input}} \times 100$$

A graph of losses and efficiency for a compound generator is shown at Fig 12.

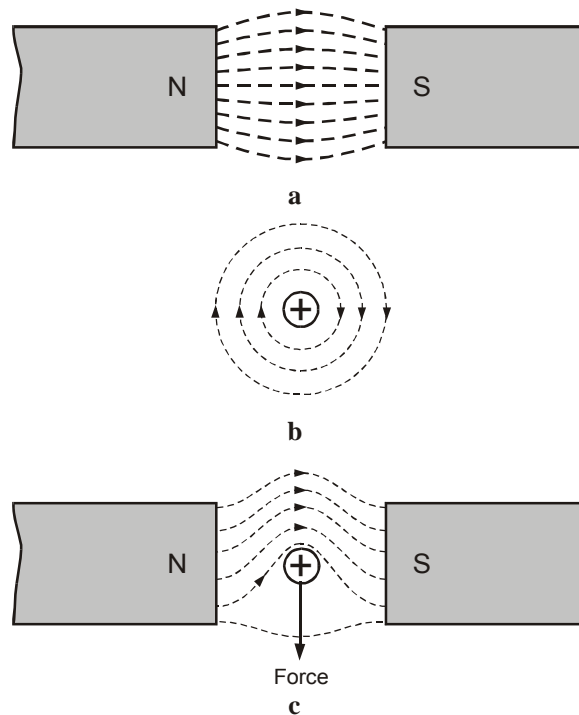
14-6 Fig 12 Losses and Efficiency of a Compound Generator



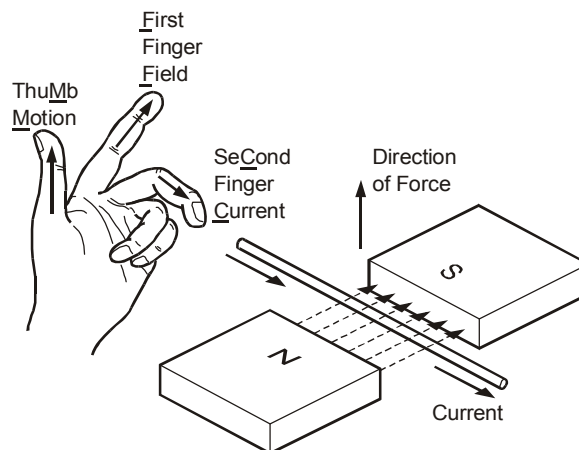
## THE DC MOTOR

### General Principles

27. An electric motor is a machine for converting electrical energy into mechanical energy, its function being the reverse of that of the generator. The conversion of the electrical energy relies on the fact that an electric current flowing through a conductor placed in a magnetic field causes the wire to move (if free to do so). The lines of flux of the two fields interact as shown in Fig 13 and, in this case, a force is created downward on the conductor.

**14-6 Fig 13 Force on a Current Carrying Conductor****Fleming's Left-hand Rule**

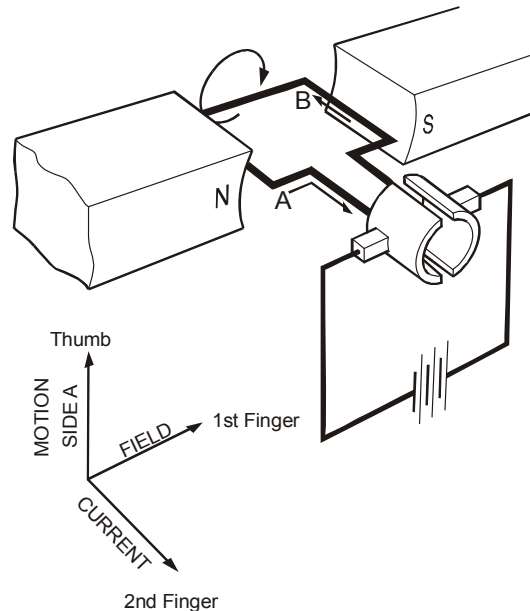
28. The direction in which a current carrying conductor will move when placed in a magnetic field can be determined from Fleming's Left-hand Rule (see Fig 14). If the first finger, the second finger, and the thumb of the LEFT hand are held at right angles to each other, then with the **F**irst finger pointing in the direction of the **F**ield (N to S), and the **se**Cond finger in the direction of the **C**urrent in the conductor, the **thu**Mb will indicate the direction in which the conductor tends to **M**ove.

**14-6 Fig 14 Fleming's Left-hand Rule****The Simple Motor**

29. A motor depends for its operation on the force exerted upon current-bearing conductors situated in a magnetic field. Consider a simple permanent magnet motor connected to a battery as shown in Fig 15. By applying Fleming's Left-hand Rule it will be seen that side A of the loop (under the N pole) tends to move upwards, while side B of the loop tends to move downwards. The forces acting on the two sides of the loop are thus cumulative in their effect, and tend to turn the loop in a clockwise direction.

30. As the sides of the loop pass through the neutral position, the commutator reverses the connections of the supply to the loop, and the current in the loop is consequently reversed. It follows that the force acting on side A will now be downwards, and on side B upwards. The mechanical force on the loop is thus continued in the original direction, and rotation continues so long as the supply is connected.

**14-6 Fig 15 DC Motor Principle**



### The Practical Motor

31. A single-loop motor is of little use and improvements like those made to a simple generator described in paras 11, 12 and 13 are incorporated. The direction of rotation of the motor depends upon the direction of the current in the armature coils and also upon the direction of the magnetic field. If one of these is reversed the direction of rotation is reversed; reversing both together has no effect.

### The Speed of a Motor

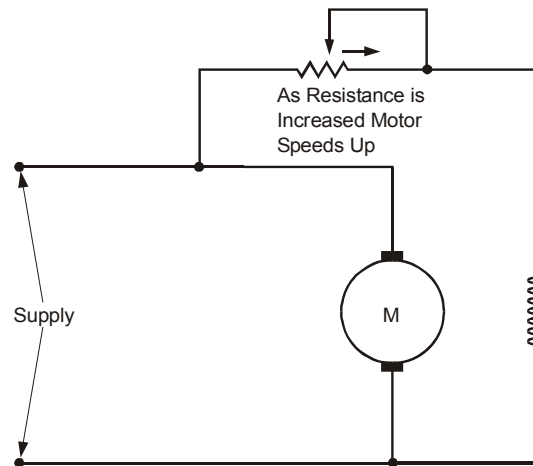
32. **The Back emf.** When the armature in the motor rotates an emf is induced in the conductors by generator action. The direction of the emf is such as to oppose the motion or other cause producing it (Lenz's law), that is it opposes the supply voltage. This emf (called the back emf) is proportional to the field strength and the speed of rotation of the armature, but can never be as great as the supply input voltage.

33. **Speed with Variable Load.** The back emf developed in the armature of a DC motor when it is running determines the current in the armature, and makes the motor a self regulating machine in which speed and armature current are automatically adjusted to the mechanical load. At small values of load the shaft torque exceeds the load torque; the armature therefore accelerates and gives rise to a large back emf. The back emf cuts down the armature current, thus reducing the shaft torque, until eventually a state of balance between the two torques is obtained and the speed is stabilized. With increasing load the load torque is increased, exceeding the shaft torque and causing a fall in armature speed. Reduced armature speed results in reduced back emf and increased armature current; the increase in armature current produces an increase in shaft torque restoring torque balance and stabilizing the speed again. The variation of speed with armature current (ie with mechanical load) is known as the speed characteristic of the motor.

34. **Control of Motor Speed.** Assuming constant load, there are two methods commonly used to vary the speed of a DC motor:

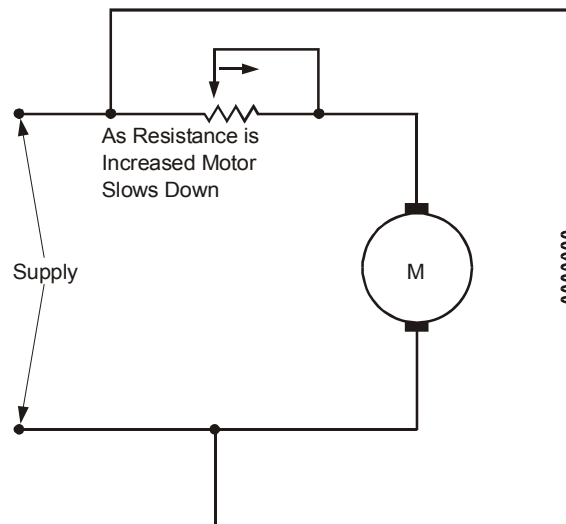
a. **Field Control** (Fig 16). By weakening the main flux of a motor the back emf is reduced, increasing the effective voltage and the armature current. The increased armature current gives rise to an increased shaft torque, causing the motor to accelerate until the back emf, rising with increased speed, restricts the armature current and shaft torque to restore the balance of shaft and load torques. At this point the speed of the motor will stabilize. Conversely, an increase in field strength will cause a reduction in speed. The minimum speed will be obtained with full field excitation, this being known as the normal speed of the machine.

14-6 Fig 16 Speed Varied by Controlling Field Strength



b. **Armature Control** (Fig 17). By reducing the voltage across the armature of a motor the effective voltage is reduced, with a corresponding reduction in armature current and shaft torque. The excess of load torque over shaft torque causes the motor to slow down to a point where the reduced back emf permits sufficient armature current to produce a state of balance between the two torques. At this point the speed of the motor will stabilize.

14-6 Fig 17 Speed Varied by Controlling Armature Current



## Types of Motors

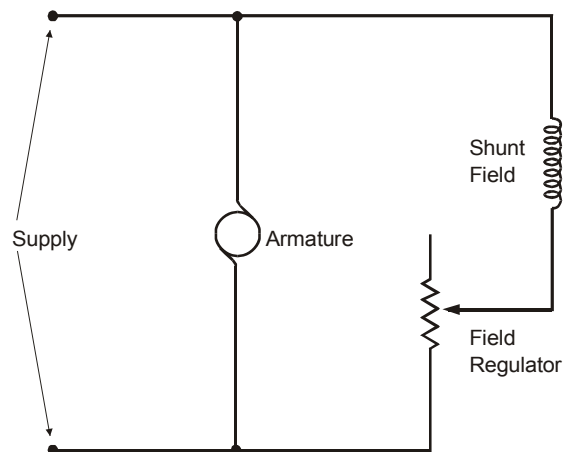
35. DC motors are classified according to the method by which the field is excited:

- a. The majority of motors are comparable to self-excited generators, ie the armature winding and field winding are supplied from a common source. Their speed and load characteristics vary according to the method of connecting the field winding to the armature, and as a class, they are capable of fulfilling most requirements.
- b. Separately-excited motors are used only for special purposes where the more normal types are unsuitable.
- c. Permanent magnet motors are employed for certain special purposes, eg in small control systems.

Only the self-excited types will be described in detail.

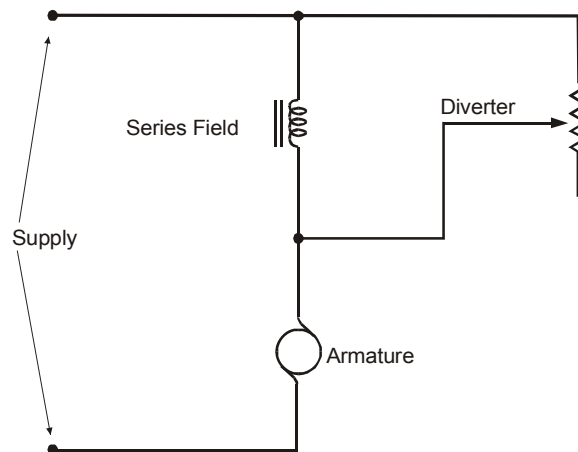
36. **Shunt-wound Motors.** The field winding of the shunt-wound motor is connected in parallel with the armature (Fig 18); it is thus directly across the supply and must be of relatively high resistance to restrict the current through it. The speed of the motor drops only slightly as the load increases and can be considered constant for many applications. The torque develops proportionally with armature current. The initial torque is small because of the restricted armature current, and shunt motors should be started on light load or no load. Its constant speed makes this type of motor suitable for lathes, drills, and light machine tools.

14-6 Fig 18 A Shunt-wound Motor



37. **Series-wound Motors.** The armature and the field winding of the series-wound motor are in series with each other across the supply (Fig 19). Thus the currents in each of them are the same or (as in Fig 19) proportional to each other. The speed decreases with increase in load, and these motors should always be started under load to prevent over-speeding at start-up. The torque increases rapidly as the load is increased, the starting torque being high. Due to its high starting torque a series-wound motor is used for engine starting and traction work.

38. **Compound-wound Motors.** By arranging for part of the field winding to be in series with the armature and part in parallel with it, the large starting torque of the series motor and the almost constant speed of the shunt can be reproduced in the compound-wound motor.

**14-6 Fig 19 A Series-wound Motor**

### Motor Starters

39. Due to the low resistance of the armature, extremely high currents would be experienced in the armature until its movement caused sufficient back emf to limit the current to an acceptable level. To prevent the high current burning out the armature windings, a starter resistance is inserted in series with them. The resistance is then progressively reduced as the back emf increases. Starter resistors are usually used with heavy motors; the armatures of small motors have sufficiently high resistance to limit the current.

### Rotary Transformers

40. The rotary transformer is a combination of a DC motor and generator. It consists of a single field system and a single armature, on which are wound the motor and generator windings. A DC input is transformed into a different DC output, a typical example being an input of 24V 7A and an output of 1,200V 100 mA. The power output is always less than the power input.



**CHAPTER 7 - ALTERNATING CURRENT THEORY**

<b>Contents</b>	<b>Page</b>
Introduction .....	2
AC WAVEFORMS .....	2
Terms Used .....	2
Waveform .....	3
Square Waveform.....	4
AC VALUES .....	5
General .....	5
Instantaneous Value.....	6
Peak Value .....	6
Average or Mean Value.....	6
Effective or Root-Mean-Square (rms) Value .....	6
REPRESENTATION OF SINUSOIDAL WAVEFORM .....	7
General .....	7
Graphical Representation.....	7
Trigonometrical Representation .....	9
Vectorial Representation .....	10
IDEAL COMPONENTS IN CIRCUITS.....	12
Introduction .....	12
Pure Resistive Circuits .....	12
Pure Inductive Circuits .....	14
Pure Capacitive Circuits .....	17
Summary .....	20

**Table of Figures**

14-7 Fig 1 Alternating Voltage .....	3
14-7 Fig 2 AC Waveforms .....	3
14-7 Fig 3 AC Superimposed on DC.....	4
14-7 Fig 4 Composition of a Square Wave .....	5
14-7 Fig 5 AC Values .....	5
14-7 Fig 6 Effective or rms Value .....	6
14-7 Fig 7 Dependence of emf on $\sin \theta$ .....	7
14-7 Fig 8 Plotting a Sine Wave.....	8
14-7 Fig 9 Phase Difference by Graphs .....	9
14-7 Fig 10 Vector Representation of AC .....	10
14-7 Fig 11 Vector Convention.....	10
14-7 Fig 12 Phase Difference by Graphs and Vectors .....	11
14-7 Fig 13 Phase Difference by Vector .....	11
14-7 Fig 14 Reference Vectors .....	11
14-7 Fig 15 Resultant of Two alternating Reference Vectors .....	12
14-7 Fig 16 A Resistive AC Circuit.....	12
14-7 Fig 17 Power Waveform - Resistive AC Circuit .....	13
14-7 Fig 18 Rate of Change of Current - Inductive AC Circuit.....	14
14-7 Fig 19 Phase Relationships in an Inductive AC Circuit.....	15
14-7 Fig 20 Vector Relationships in an Inductive AC Circuit .....	16

14-7 Fig 21 Variation of Inductive Reactance with Frequency .....	16
14-7 Fig 22 Power Curve of a Purely Inductive Circuit .....	17
14-7 Fig 23 Phases of Current and Voltage in a Pure Inductance .....	17
14-7 Fig 24 Phase Relationship in a Capacitance .....	18
14-7 Fig 25 Variation of $X_C$ with Frequency .....	19
14-7 Fig 26 Phases of Current and Voltage in a Pure Capacitance .....	20

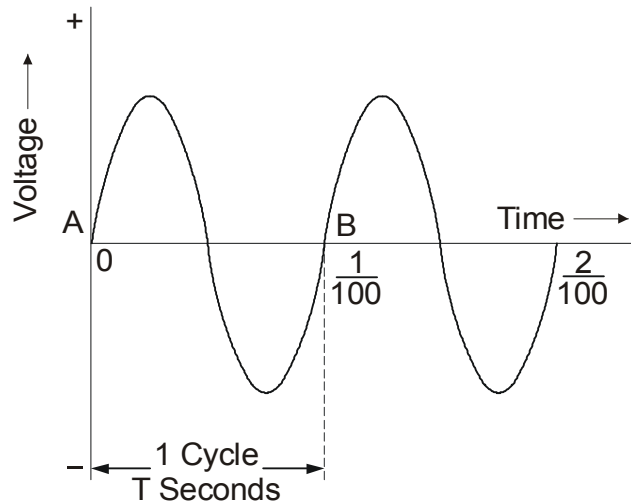
## Introduction

1. It was stated earlier that a direct current is one in which the electron flow is in one direction only through the circuit, whereas an alternating current is one in which the electrons flow first in one direction then the other, changing direction at regular, very short time intervals.
2. The national electrical power supply network delivers power, the mains, to consumers as an AC. Two advantages of AC over DC in this case are:
  - a. With AC, voltages can be readily stepped up or stepped down by transformers (which will be discussed later), whereas with DC conversion from one voltage to another is cumbersome and expensive.
  - b. With AC, the loss of power in transmission is minimized. The power carried by a supply line is  $EI$ ; the power lost in transmission due to the resistance of the cables is  $I^2R$ . By using a high voltage and a small current high power can be carried with little loss.
3. For the mains supply, the AC voltage generated at the power station is stepped up to a high value for transmission over the national grid, and then stepped down at sub-stations to the nominal mains supply voltage (usually 240 V) for domestic lighting, heating, etc.

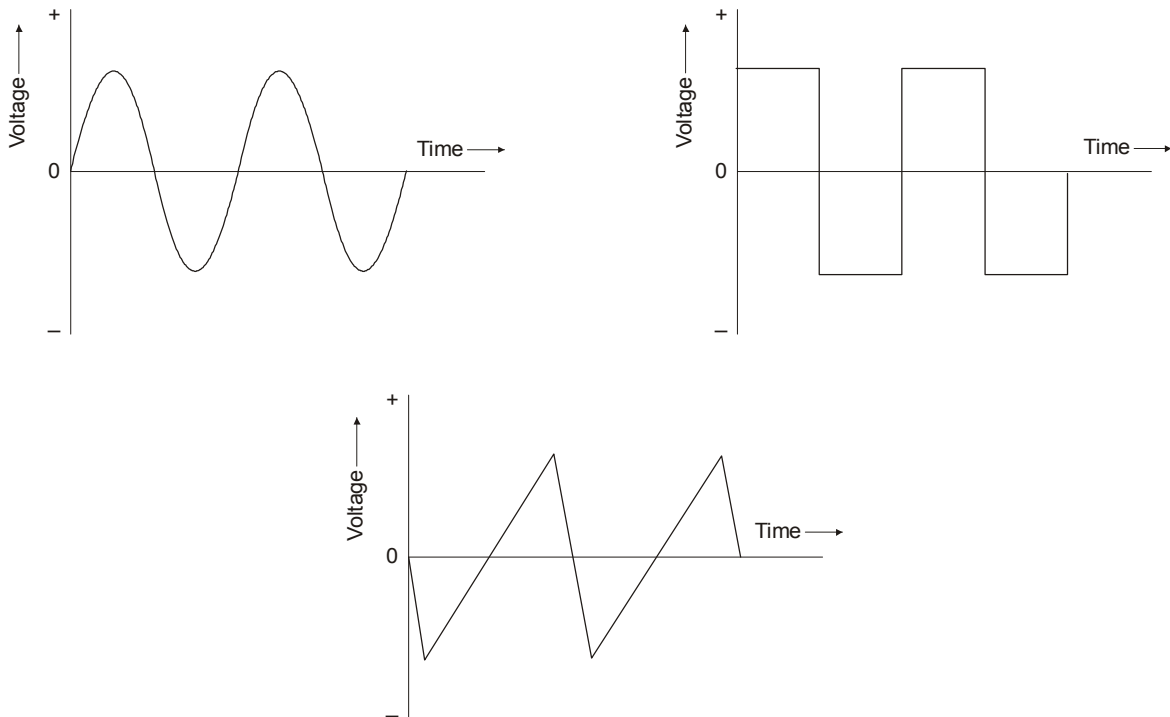
## AC WAVEFORMS

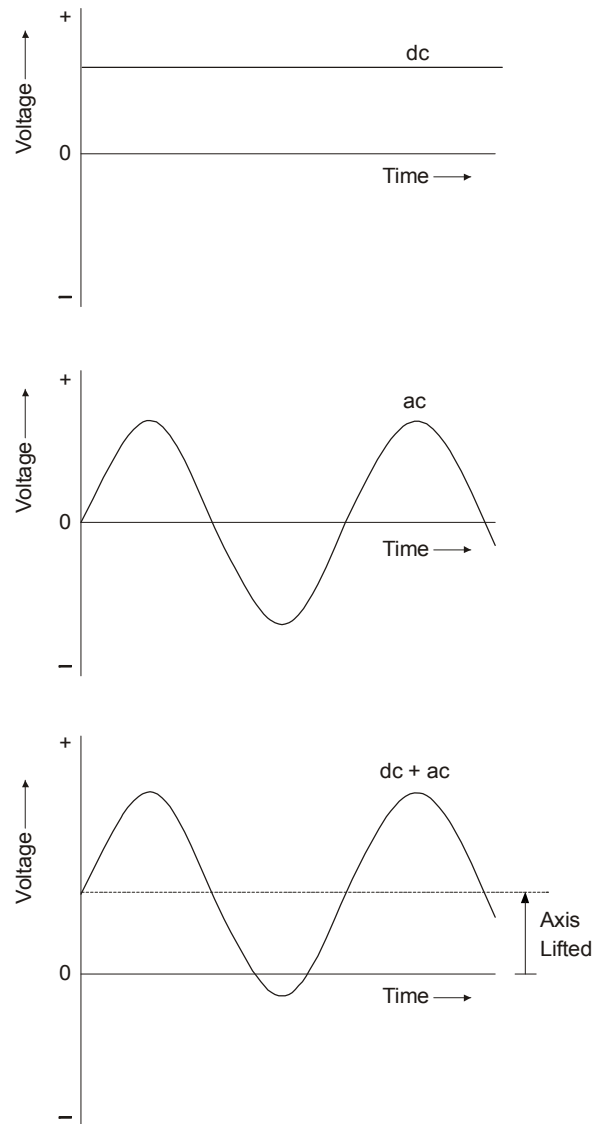
### Terms Used

4. The production of an alternating emf by a simple generator was discussed in Volume 14, Chapter 6. Fig 1 is a graph of an alternating voltage, showing voltage against time. Some of the terms associated with such a waveform are:
  - a. **Cycle.** A complete sequence of positive and negative values, such as that from A to B.
  - b. **Period (or Periodic Time).** The duration of one cycle, denoted by the symbol  $T$ . In Fig 1,  $T = 1/100$  second.
  - c. **Frequency.** The number of cycles completed in one second, denoted by the symbol  $f$ . It will be apparent that  $f=1/T$  cycles per second. In Fig 1, since  $T=1/100$  sec,  $f = 100$  cycles per second, i.e. 100 hertz (Hz). Low frequency AC currents are used in the supply of electrical power. The AC mains supply in Britain has a frequency of 50 hertz; aircraft generally use 400 hertz. In radio, AC currents with frequencies ranging up to hundreds of megahertz are used.

**14-7 Fig 1 Alternating Voltage****Waveform**

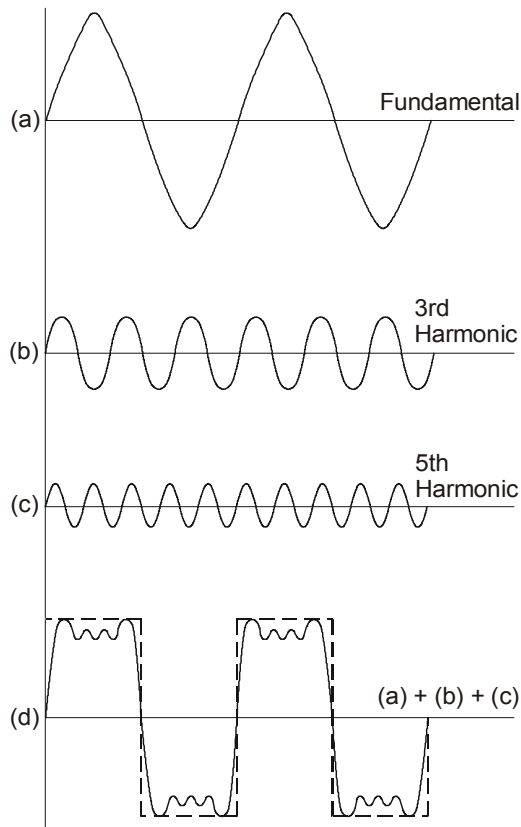
5. The shape of the AC voltage or current graph is known as the waveform; in Fig 1, it is a sine curve. Fig 2 shows three voltages with the same frequency but different waveforms. In some circuits, AC and DC are both present, and add together in such a way that the AC is superimposed on the DC. The result is that the AC axis shifts from the zero line (Fig 3). Note that if the resulting voltage never falls below zero, the current is called a pulsating DC current.

**14-7 Fig 2 AC Waveforms**

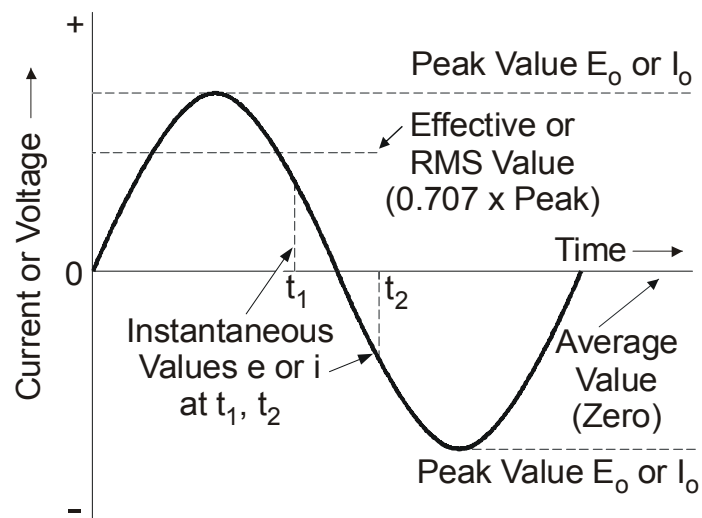
**14-7 Fig 3 AC Superimposed on DC**

### Square Waveform

6. It is of interest here to note how a square wave is produced. Fourier's theorem states that any recurrent wave-form of frequency  $f$  can be resolved into the sum of a number of sinusoidal wave-forms having frequencies  $f$ ,  $2f$ ,  $3f$  etc; the number of sine waves may be finite or infinite. The frequency  $f$  is known as the fundamental or first harmonic, and frequencies  $2f$ ,  $3f$ ,  $4f$ , etc as the second, third, fourth etc harmonics. The amplitude of a harmonic is inversely proportional to its frequency. The square wave consists of a fundamental sine wave and all its odd harmonics up to infinity. Fig 4 shows the result of taking harmonics up to the fifth, and it will be seen that this gives a good approximation to a square wave.

**14-7 Fig 4 Composition of a Square Wave****AC VALUES****General**

7. Several things can be meant by the value of an alternating quantity such as an AC current; the various terms used to make it clear which meaning is intended are defined in the following paragraphs and illustrated at Fig 5.

**14-7 Fig 5 AC Values**

### Instantaneous Value

8. The value of an alternating quantity is continuously changing, its instantaneous value being that at any given instant of time. Instantaneous current is denoted by the symbol 'i', and instantaneous voltage by 'e' (Fig 5).

### Peak Value

9. The maximum value of an alternating quantity (either positive or negative) reached during a cycle is called the peak value. Peak current is denoted by  $I_0$  (or  $I_{\max}$ ) and peak voltage by  $E_0$  (or  $E_{\max}$ ).

### Average or Mean Value

10. The average value of an alternating quantity over a number of complete cycles is zero. In Fig 5, the area between the current curve and the time axis represents the quantity of electricity that has passed during the interval concerned. As the curve is symmetrical about the time axis (provided that as in Fig 5 it represents AC with no DC present), the quantity of electricity that passes one way during the first half exactly equals that passing the other way during the second half of any cycle. The net transfer of electricity, and hence the average current, is zero.

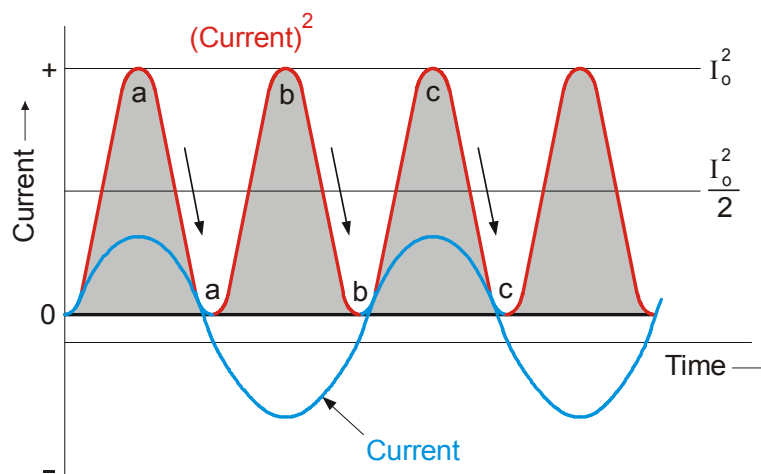
### Effective or Root-Mean-Square (rms) Value

11. The heating effect of an electric current is independent of the direction of the current, and the effective value of an alternating current is 1 ampere if it has the same heating effect as an unvarying direct current of 1 ampere.

12. It was shown in Volume 14, Chapter 1, when considering power, that the rate of dissipation of energy as heat in a resistor depended upon the square of the current or voltage, ie  $P = I^2 R = V^2/R$  watts. This applies also to AC. In Fig 6, the square of the sinusoidal current is plotted against time, giving a curve which is always positive and is symmetrical about the halfway line  $\frac{I_0^2}{2}$ . The energy

dissipated over any time interval is proportional to the shaded area beneath the curve of  $(\text{current})^2$ . In this graph, crest a will fit into trough a, crest b into trough b, and so on, so that the area beneath the curve of  $(\text{current})^2$  equal to the area beneath the horizontal line drawn midway between the square of the peak current and the time axis. Thus the heating effect of the current plotted at Fig 6 is the same as that of a direct current of value  $\sqrt{(I_0^2/2)}$ , or  $I_0/\sqrt{2}$ .

14-7 Fig 6 Effective or rms Value



13. This value is known as the effective value (since it is that value which has the same heating effect as an equivalent direct current), or as the root-mean-square (rms) value (since it is obtained by finding the mean value of the square of the current and then taking the square root). If  $I_{\text{rms}}$  is the rms value of alternating current in a circuit, it is related to its peak value by the expression:

$$I_{\text{rms}} = \frac{I_0}{\sqrt{2}} = 0.707 I_0 \text{ amps}$$

$$\text{Also, } I_0 = I\sqrt{2} = 1.414 I \text{ (amps)}$$

14. Alternating currents and voltages are usually measured by their rms values. For instance, the normal AC mains supply has a voltage of 240 V; this is an rms value, the peak voltage being  $240\sqrt{2}$  which equals 340 V. Ammeters and voltmeters for AC are normally calibrated in rms values.

## REPRESENTATION OF SINUSOIDAL WAVEFORM

### General

15. Alternating currents and voltages have so far been represented by graphs. In the following paragraphs this method of representing sinusoidal wave-forms will be examined in more detail, and brief mention will be made of two other methods often encountered: representation by trigonometrical equations and by vectors.

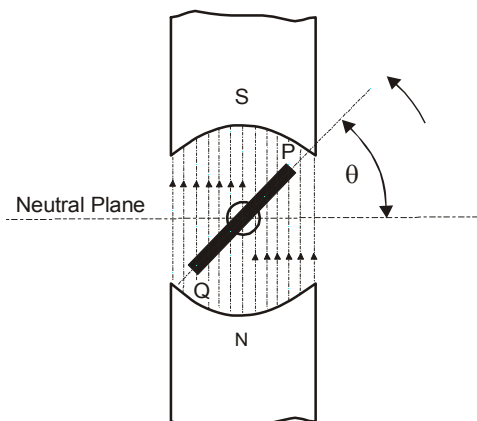
### Graphical Representation

16. Consider a conductor loop (PQ) rotating in an anti-clockwise direction at constant speed in a uniform magnetic field, as shown in Fig 7. When the loop is in the neutral plane, at right angles to the magnetic field ( $\theta = 0^\circ$ ), the emf induced in the loop is zero. When the loop is at right angles to the neutral plane ( $\theta = 90^\circ$ ) the emf is a maximum ( $E_0$ ), its value depending on the speed of rotation and on the flux density. At intermediate points, the induced emf is between zero and its maximum value, and depends on the angle which the loop makes to the neutral plane. The emf induced in the loop at any instant is proportional to the sine of the angle ( $\theta$ ) through which the loop has rotated from the neutral plane, i.e.:

$$e = E_0 \sin \theta \text{ (volts).}$$

For example, when the loop is in the neutral plane,  $\theta = 0^\circ$ ,  $\sin \theta = 0$ , and  $e = E_0 \sin \theta = \text{zero}$ . When the loop is at right angles to the neutral plane,  $\theta = 90^\circ$ ,  $\sin \theta = 1$ , and  $e = E_0 \sin \theta = E_0$  volts.

**14-7 Fig 7 Dependence of emf on Sin  $\theta$**



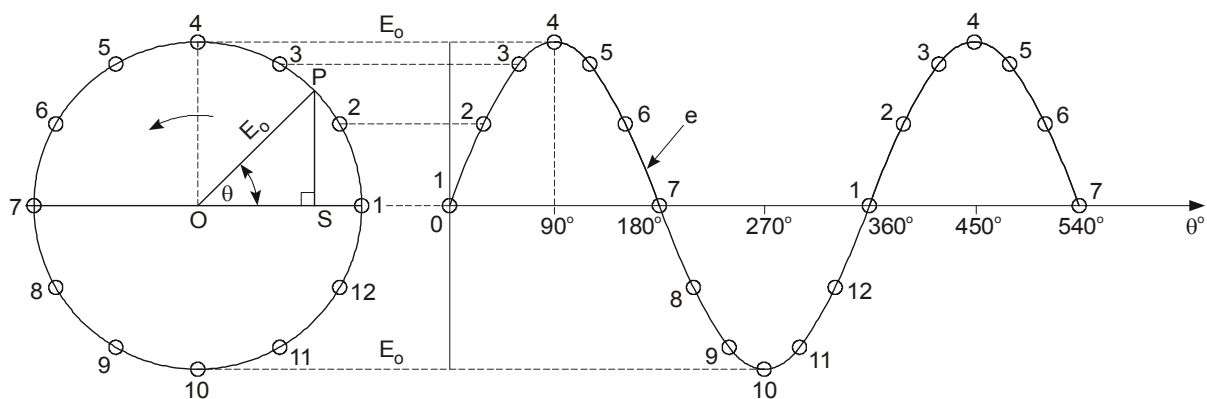
17. To plot a curve showing the instantaneous value of the emf for all values of the angle  $\theta$ , consider Fig 8. The line OP is assumed to rotate about the point O in an anti-clockwise direction, its length representing the maximum value of the emf ( $E_0$ ) to any convenient scale. A horizontal line is drawn through the centre of rotation (O) and, along this, a scale of degrees of rotation is marked. Now:

$$\sin \theta = \frac{PS}{OP}$$

$$\therefore PS = OP \sin \theta = E_0 \sin \theta$$

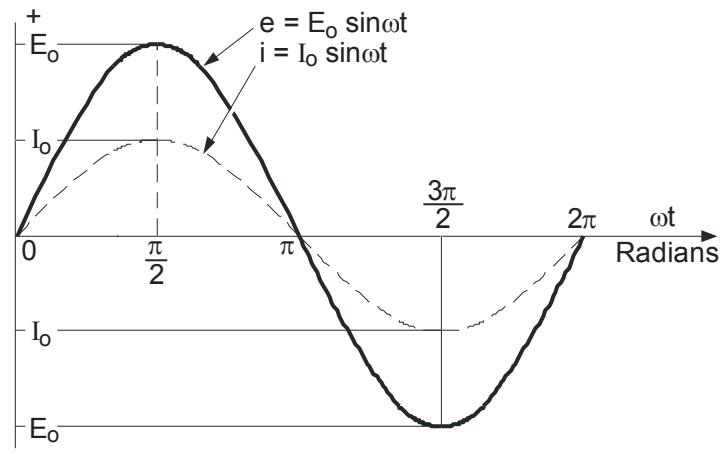
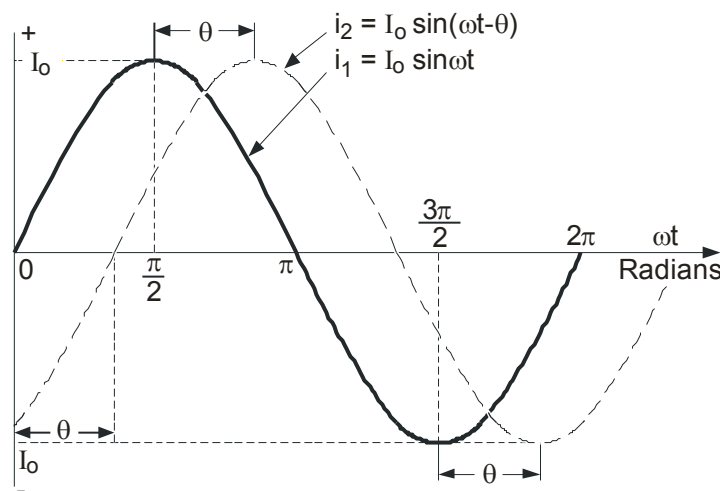
Thus, the vertical line PS represents, to scale, the instantaneous emf 'e' =  $E_0 \sin \theta$ . As OP rotates, the length of the line PS varies, and plotting this against the various values of the angle  $\theta$  on the horizontal axis, the graph of the instantaneous emf is obtained. This curve is termed a sine curve, and any quantity which varies in this manner is said to have a sinusoidal waveform.

**14-7 Fig 8 Plotting a Sine Wave**



18. **Phase Difference.** When two alternating quantities of the same frequency pass through corresponding points in a cycle at the same instant of time they are said to be in phase with each other (Fig 9a). If they pass through corresponding points at different instants of time, there is a phase difference between them and one is said to be leading or lagging the other by a certain phase angle. For example, in Fig 9b,  $i_1$  is leading  $i_2$  by  $\theta$  radians (or  $i_2$  is lagging by  $\theta$  radians). Thus,  $i_1$  reaches its maximum value  $\theta$  radians before  $i_2$ .



**14-7 Fig 9 Phase Difference by Graphs****a In Phase****b Out of Phase****Trigonometrical Representation**

19. **Angular Velocity.** Normally the size of angles is expressed in degrees. When dealing with rotation another unit of measurement, the radian, is used. A radian is the angle subtended at the centre of a circle by an arc of the circumference equal in length to the radius of that circle. As the total length of the circumference is  $2\pi r$  and a complete revolution is  $360^\circ$ .

$$360^\circ = 2\pi \text{ radians, or } 1 \text{ radian} = 57.3^\circ$$

20. **Instantaneous Values of e and i.** Using radians, expressions for the instantaneous values of e and i, which are related to time, can be derived. In one revolution, the loop passes through  $360^\circ$  or  $2\pi$  radians. If it rotates at f revolutions per second it passes through  $2\pi f$  radians per second; this is termed the angular velocity of the loop and is denoted by the Greek letter  $\omega$ . Thus:

$$\text{angular velocity} = \omega \text{ (radians per sec)}$$

After an interval of t seconds from the commencement of rotation the loop has rotated through an angle  $\theta$  equal to  $2\pi ft = \omega t$  radians, and the emf at this instant is:

$$e = E_0 \sin \theta$$

$$= E_0 \sin \omega t \text{ (volts)}$$

Similarly, with an alternating current in a circuit, the value of the current at any instant 't' is:

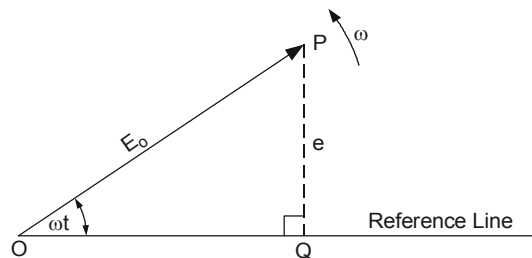
$$i = I_0 \sin \omega t \text{ (amps)}$$

**21. Phase Difference.** The information in the graph of Fig 9b can also be conveyed by trigonometrical equations. For instance, the pair of equations  $i_1 = I_0 \sin \omega t$  and  $i_2 = I_0 \sin (\omega t - \theta)$  indicate that the two currents  $i_1$  and  $i_2$  have the same amplitude  $I_0$  and the same angular velocity,  $\omega$ , but  $i_2$  is lagging  $i_1$  by  $\theta$  radians.

### Vectorial Representation

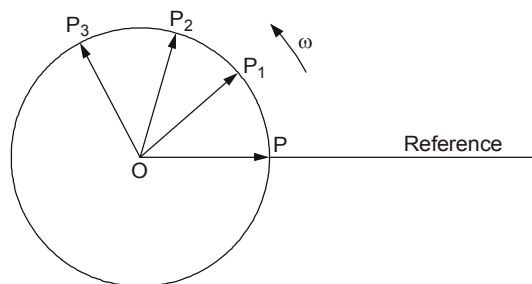
22. The trigonometrical representation of voltages and currents, described in para 20, can be depicted by vectors. In the special application of vectors to alternating quantities, the vector is assumed to be fixed at its origin but free to rotate at the frequency of the alternations, the length of the vector being equal to the peak value of the alternating quantity. For example, consider a voltage  $e = E_0 \sin \omega t$ . This voltage may be represented by a vector of length corresponding to  $E_0$  rotating in an anti-clockwise direction with reference to a datum or reference line, with angular velocity  $\omega$  radians per second. This is shown in Fig 10 where, after time 't' seconds, the vector has rotated through an angle of  $\omega t$  radians. The instantaneous value of  $e$  is shown to the same scale as the vector  $E_0$  by the line PQ (see para 17).

14-7 Fig 10 Vector Representation of AC



23. While the voltage  $E_0$  throughout one cycle could be represented by a complete series of lines  $OP$ ,  $OP_1$ ,  $OP_2$ , etc (Fig 11), it is usually sufficient to show one position of the vector, for its rotation with angular velocity  $\omega$  is understood. It is usual to take the instant  $t = 0$  for the voltage  $e = E_0 \sin \omega t$  as the reference line.

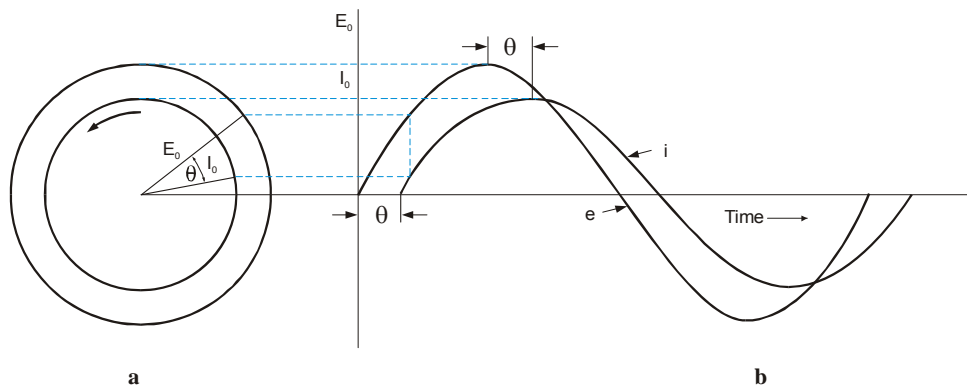
14-7 Fig 11 Vector Convention



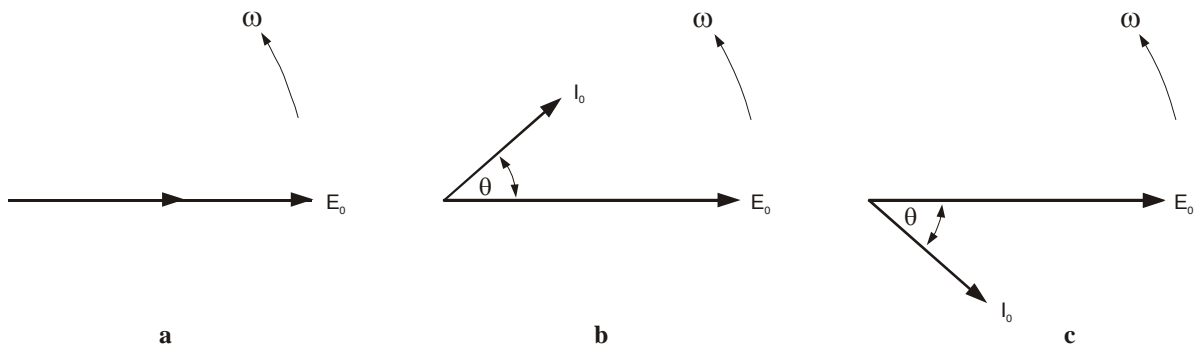
24. Figs 12a and 12b both show the phase difference between a current and its related voltage, but the vector representation is more simple to construct and use. In Fig 13, let  $E_0$  represent the peak value of an alternating voltage, and  $I_0$  the peak value of the current in a circuit. Then, if the voltage and current are in phase,  $E_0$  and  $I_0$  may be represented by vectors which are coincident in direction and rotate at equal angular velocity  $\omega$  radians per second.  $E_0$  and  $I_0$  are therefore represented as in

Fig 13a. In practice  $I_0$  may be alternatively in phase with, or lead, or lag the voltage  $E_0$ . In Fig 13b,  $I_0$  is shown as leading by angle  $\theta$ , and in Fig 13c as lagging by angle  $\theta$  with reference to  $E_0$ .

14-7 Fig 12 Phase Difference by Graphs and Vectors

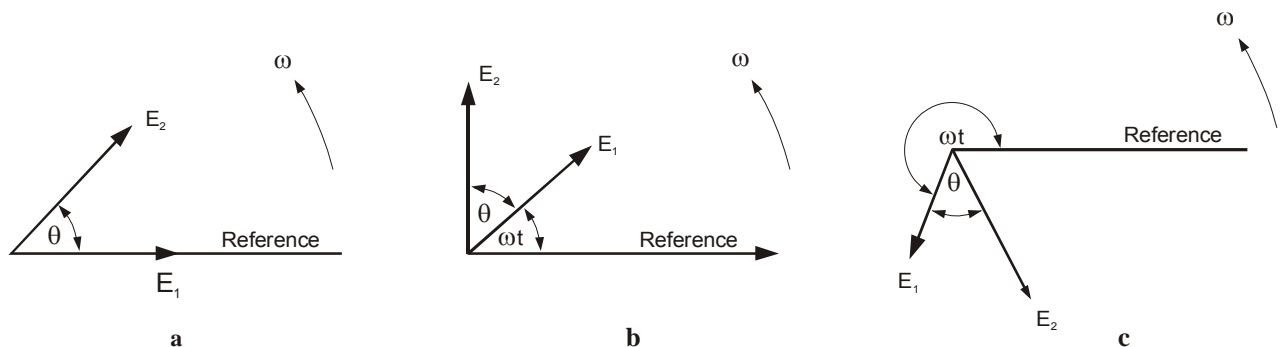


14-7 Fig 13 Phase Difference by Vector



25. **Reference Vectors.** Consider two voltages  $e_1 = E_1 \sin \omega t$  and  $e_2 = E_2 \sin (\omega t + \theta)$ . These two voltages can be represented by vectors of length  $E_1$  and  $E_2$  respectively, rotating with the same angular velocity  $\omega$  radians per second and displaced from each other by an angle  $\theta$ . At the instant  $t = 0$ ,  $E_1$  is lying along the reference vector and  $E_2$  is leading by an angle  $\theta$ , as shown in Fig 14a. Since both vectors are rotating at the same angular velocity, they maintain the same relative positions as shown in Figs 14b and 14c. In AC problems, it is sufficient to consider vectors in one position only and it is usual to take the instant  $t = 0$  as depicted in Fig 14a when one of the vectors lies along the reference line.

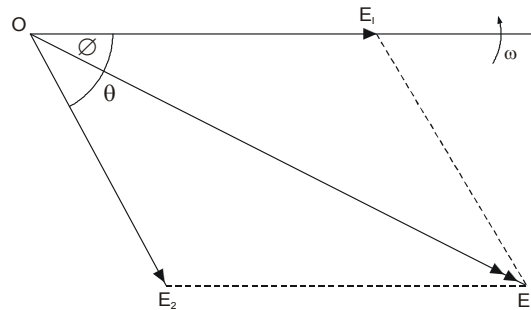
14-7 Fig 14 Reference Vectors



26. **Resultant of Two Voltages.** A single voltage, equivalent to any two alternating voltages acting across the same circuit at the same time, can be found by the parallelogram rule for finding the resultant of two vectors. Thus, in Fig 15,  $E_1$  and  $E_2$  are the vectors representing two voltages

$e_1 = E_1 \sin \omega t$  volts and  $e_2 = E_2 \sin (\omega t - \theta)$  volts respectively. The resultant voltage is represented by the diagonal of the parallelogram drawn outwards from O, ie the vector  $E_R$ . This resultant vector rotates at the same angular velocity  $\omega$  as  $E_1$  and  $E_2$  and with reference to  $E_1$  lags by an angle  $\phi$ . The resultant voltage is, therefore,  $e_R = E_R \sin (\omega t - \phi)$  volts.

**14-7 Fig 15 Resultant of Two alternating Reference Vectors**



## IDEAL COMPONENTS IN CIRCUITS

### Introduction

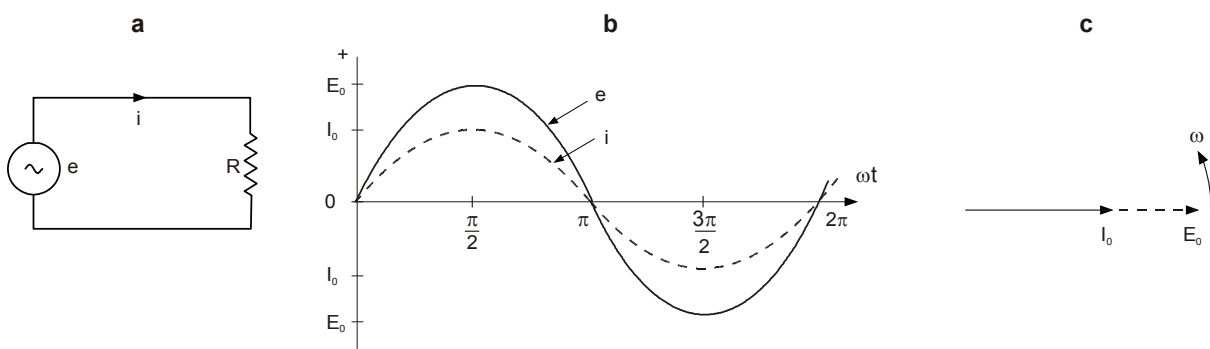
27. In the next chapter, the combined effects of resistance, inductance, and capacitance in AC circuits are examined in some detail. In this chapter a simpler approach is adopted, the assumption being made that resistance, inductance and capacitance exist separately; as has already been stated in discussing DC circuits this is an idealized approach, inductors for example having some resistance, and wire-wound resistors some inductance.

28. When purely resistive components are used in AC circuits, Ohm's Law, Kirchoff's Laws, and the usual circuit rules apply exactly as in DC circuits, noting that in general cases rms values of current and voltage are used.

### Pure Resistive Circuits

29. When an alternating voltage is applied across a resistance (Fig 16a), the waveform of the voltage is a sine wave. As Ohm's Law applies at any instant, the current flow follows the same waveform, increasing and decreasing with the voltage and changing direction with it (Fig 16b). The voltage and current are said to be in phase (sine waves being in phase when they have the same frequency, and pass through zero together in the same direction, although they do not necessarily have the same amplitude).

**14-7 Fig 16 A Resistive AC Circuit**



30. **Current.** Since Ohm's Law applies at all times to a purely resistive circuit, the rms value of current is given by:

$$I_{\text{rms}} = \frac{E_{\text{rms}}}{R}$$

31. **Power.** The power in AC circuits is the average value of all the instantaneous values for a complete cycle. The instantaneous power is the product of  $e$  and  $i$  at that instant. If the multiplying process is carried out over a complete cycle, the power curve at Fig 17 is obtained (note that as  $e$  and  $i$  are both negative together, their product is always positive). The total power over one cycle is represented by the area beneath the power curve; the same area lies beneath the straight line AB drawn midway between maximum and minimum values of power. The average power over a complete cycle is thus half the peak power. Note that:

a. The power waveform has twice the frequency of the supply.

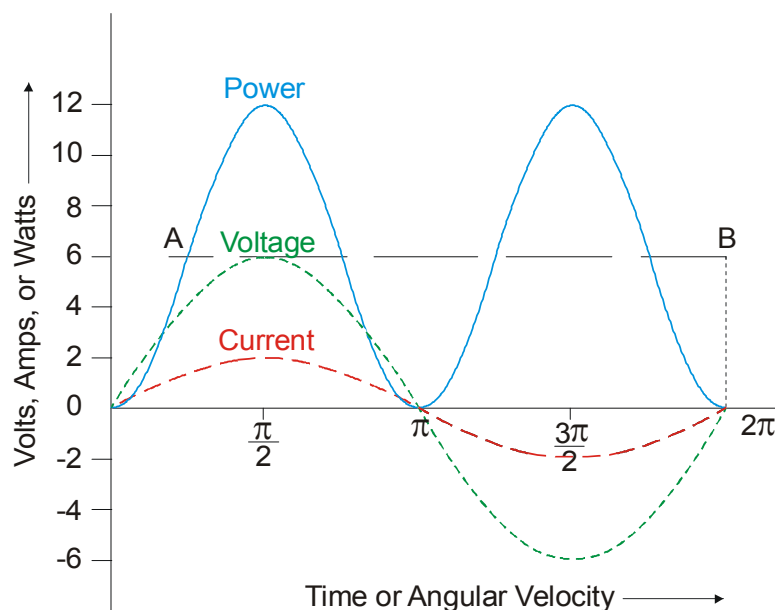
b. Maximum power ( $P_0$ ) =  $E_0 \times I_0$

$$\begin{aligned} \text{c. Average power } P_{\text{average}} &= \frac{E_0 \times I_0}{2} = \frac{E_0}{\sqrt{2}} \times \frac{I_0}{\sqrt{2}} \\ &= E_{\text{rms}} \times I_{\text{rms}} \end{aligned}$$

d. Also, since  $E_{\text{rms}} = I_{\text{rms}} \times R$ ,

$$\begin{aligned} P_{\text{average}} &= I_{\text{rms}}^2 \times R \\ &= \frac{E_{\text{rms}}^2}{R} \end{aligned}$$

14-7 Fig 17 Power Waveform - Resistive AC Circuit



32. **Deduction.** The observations made in paras 29 and 30 can be deduced using the trigonometrical approach:

$$\text{Applied voltage (e)} = E_0 \sin \omega t \text{ volts}$$

From Ohm's Law,

$$\text{current (i)} = \frac{e}{R} = \frac{E_0 \sin \omega t}{R} \text{ amps}$$

These equations show that  $i$  is a sinusoidal current in phase with  $e$  and having the same frequency  $f = \omega/2\pi$  Hz. The peak value of  $i$  occurs when  $e = E_0$  volts (ie when  $\sin \omega t = 1$ ), thus  $E_0/R = I_0$  amps. Thus:

a. The current and voltage are in phase.

b. The ratio  $\frac{E_0}{I_0} = \frac{E_{rms}}{I_{rms}} = R$  ohms

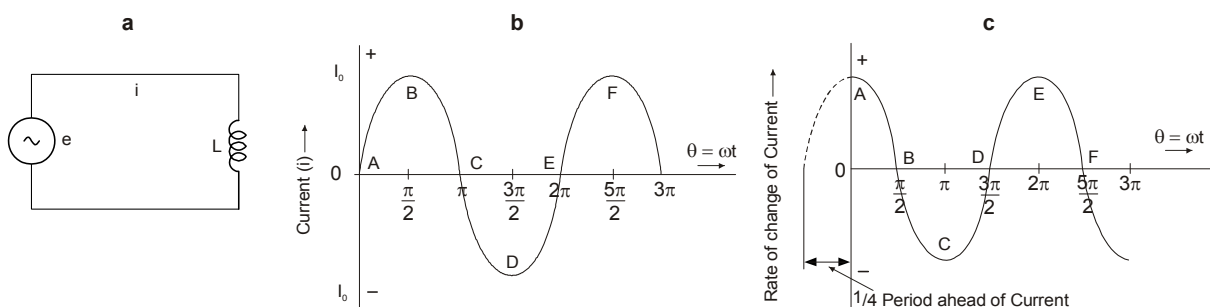
Fig 16c portrays the current and voltage vectorially.

### Pure Inductive Circuits

33. It was shown in Volume 14, Chapter 3 that a change of flux through a coil induces a voltage in it, and that a change of flux through a coil can be produced merely by varying the current in it. In steady DC circuits the current changes only when the circuit is being switched on and off; in AC circuits the current is continuously changing, and thus a voltage is induced in a coil all the time AC supply is connected. It was noted that the voltage induced (the back emf) acts in opposition to the applied voltage, thus opposing any current change.

34. The back emf induced in a coil by a changing current is given by  $e = -L di/dt$ , where  $di/dt$  is an expression meaning the rate of change of current with respect to time and  $L$  is the inductance of the coil. Fig 18b shows the curve for a sinusoidal current  $i = I_0 \sin \omega t$  which is established in the circuit of Fig 18a. The slope of the curve is horizontal at the points B, D and F, when  $\theta = \pi/2, 3\pi/2, 5\pi/2$ . Therefore, the rate of change of current is zero at these instants, giving points B, D, and F in Fig 18c. The curve has maximum and equal gradients at the points A, C and E of Fig 18b, when  $\theta = 0, \pi$  and  $2\pi$ . At these instants, the rate of change has a maximum value. At A and E, the current is positive-going and the rate of change has a maximum positive value; at C, the current is negative-going and the rate of change has a maximum negative value. Joining the points A, B, C, D, E, and F by a curve as in Fig 18c gives the curve for the rate of change of current. Such a curve is itself a sine wave leading the current waveform by  $\pi/2$  radians or  $90^\circ$ .

14-7 Fig 18 Rate of Change of Current - Inductive AC Circuit



**35. Voltage and Current Phases.** For a pure inductance the back emf is  $e = -L \, di/dt$  volts; it is in anti-phase to the rate of change of current (by virtue of the negative sign). The applied emf is, however, equal and opposite to the back emf (Kirchoff's second law) and is given by  $e = +L \, di/dt$  volts; it is, therefore, in phase with the rate of change of current. Hence, since the rate of change of current leads the current by  $\pi/2$  radians, the applied emf is a sinusoidal wave leading the current by  $\pi/2$  radians. Therefore, in a pure inductance the current and voltage are  $90^\circ$  out of phase, the current lagging the applied voltage by  $\pi/2$  radians or  $90^\circ$ . This is shown graphically in Fig 19 and vectorially in Fig 20.

**36. Inductive Reactance.** In a pure resistance the ratio of voltage to current gives the resistance  $R$ . In a pure inductance the ratio of voltage to current is:

$$\frac{E_0}{I_0} = \frac{E}{I} = \omega L = 2\pi f L \text{ (ohms)}$$

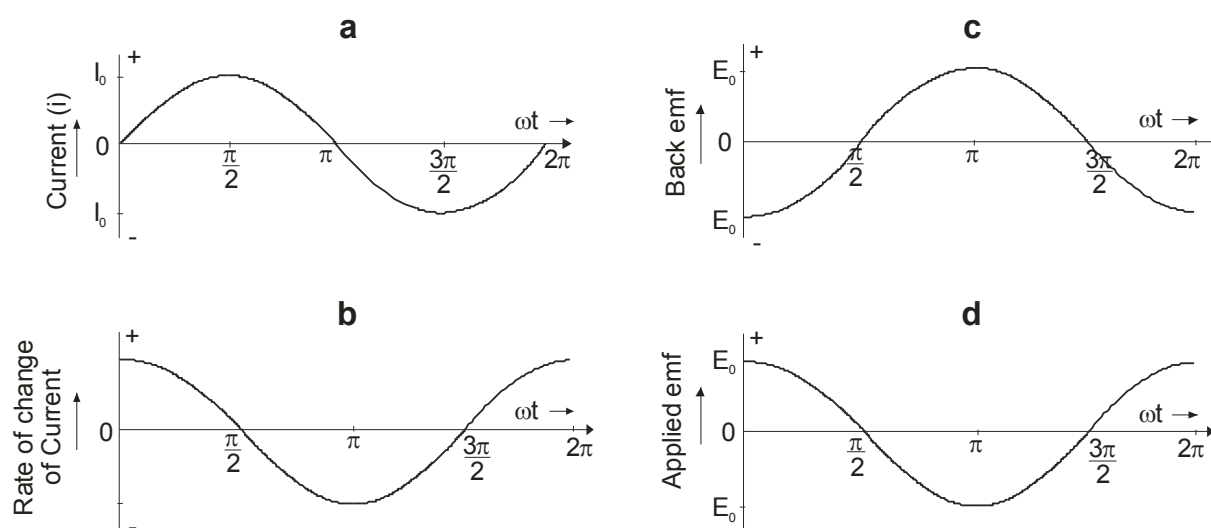
$$\text{Thus, } I = \frac{E}{2\pi f L} \text{ (amps)}$$

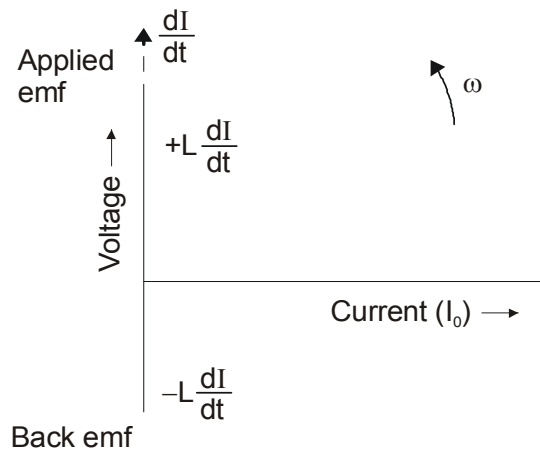
Hence, in a circuit having inductance only, the current is directly proportional to the applied voltage and inversely proportional to the frequency and the inductance. The opposition offered by pure inductance to the establishment of a current is termed the inductive reactance. The term reactance is used instead of resistance because, as can be seen from the formula, the opposition depends upon frequency. Inductive reactance is denoted by the symbol  $X_L$  and is expressed in ohms. Thus:

$$X_L = E/I = 2\pi f L,$$

where  $X_L$  = inductive reactance in ohms,  
 $f$  = frequency in hertz, and  
 $L$  = inductance in henrys.

**14-7 Fig 19 Phase Relationships in an Inductive AC Circuit**



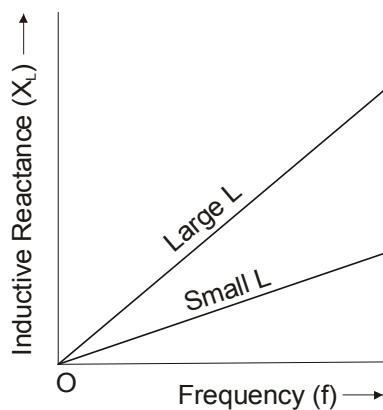
**14-7 Fig 20 Vector Relationships in an Inductive AC Circuit**

37. **The Reactance Sketch.** The reactance of an inductance is proportional to  $\omega$  and so directly proportional to the frequency  $f$ , since  $\omega = 2\pi f$ . A graph of inductive reactance against frequency for two different values of inductance is at Fig 21. This graph is known as a reactance sketch, and for a pure inductance is a straight line through the origin at O. Thus, for a 1 mH inductance:

when  $f$  is 50 Hz,  $X_L = 2\pi \times 50 \times 10^{-3} = 0.314 \Omega$

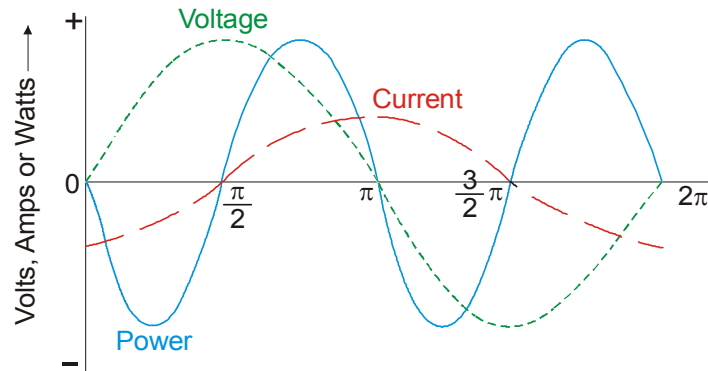
when  $f$  is 1 kHz,  $X_L = 2\pi \times 10^3 \times 10^{-3} = 6.28 \Omega$

when  $f$  is 1 MHz,  $X_L = 2\pi \times 10^6 \times 10^{-3} = 6,280 \Omega$

**14-7 Fig 21 Variation of Inductive Reactance with Frequency**

38. **Power.** The power curve for a purely inductive circuit is derived as usual by multiplying together corresponding values of current and voltage to find the instantaneous power, and plotting the products (Fig 22). It will be seen that because current and voltage are not in phase the power is sometimes positive and sometimes negative, that the frequency of the power curve is twice that of the supply, that positive power equals negative power, and that true power equals zero. Meanings can be given to the terms positive and negative power. Power is reckoned positive when energy is being taken from the generator to be stored in the magnetic field of the inductor, and negative when power is being extracted from this magnetic field and returned to the generator.

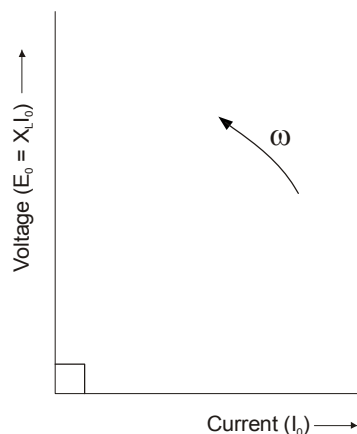


**14-7 Fig 22 Power Curve of a Purely Inductive Circuit**

39. **Conclusions.** In a purely inductive circuit:

- The current lags the applied voltage by  $90^\circ$  ( $\pi/2$  radians).
- The ratio of voltage to current gives the inductive reactance.

These two statements are illustrated by the vector diagram at Fig 23.

**14-7 Fig 23 Phases of Current and Voltage in a Pure Inductance**

### Pure Capacitive Circuits

40. In Fig 24:

- A pure capacitance is shown in Fig 24a connected across an alternating voltage supply.
- The applied voltage is  $e = E_0 \sin \omega t$ , represented graphically at Fig 24b.
- From the general expression  $Q = CV$  (where  $Q$  is the charge in coulombs,  $C$  the capacitance in farads, and  $V$  the applied emf in volts), the charge on the capacitor at the time  $t$  is given by  $q = CE_0 \sin \omega t$ , represented graphically at Fig 24c, the charge being in phase with the applied voltage.
- The current  $i$  at time  $t$  is the rate of change of charge ( $dq/dt$ ) at that instant. Using the arguments of para 34, the curve for the rate of change of charge can be shown to be itself sinusoidal, leading the curve for charge (and therefore the curve for applied voltage) by  $\pi/2$  radians or  $90^\circ$  (Fig 24d).

41. The conclusion that the curve for voltage lags behind the curve for current is expected if one considers that when the supply voltage is applied to a capacitive circuit the current rises to a maximum value immediately. However, as the voltage builds up on the plates of a capacitor so the opposition (reactance) increases, thus delaying the voltage rise.

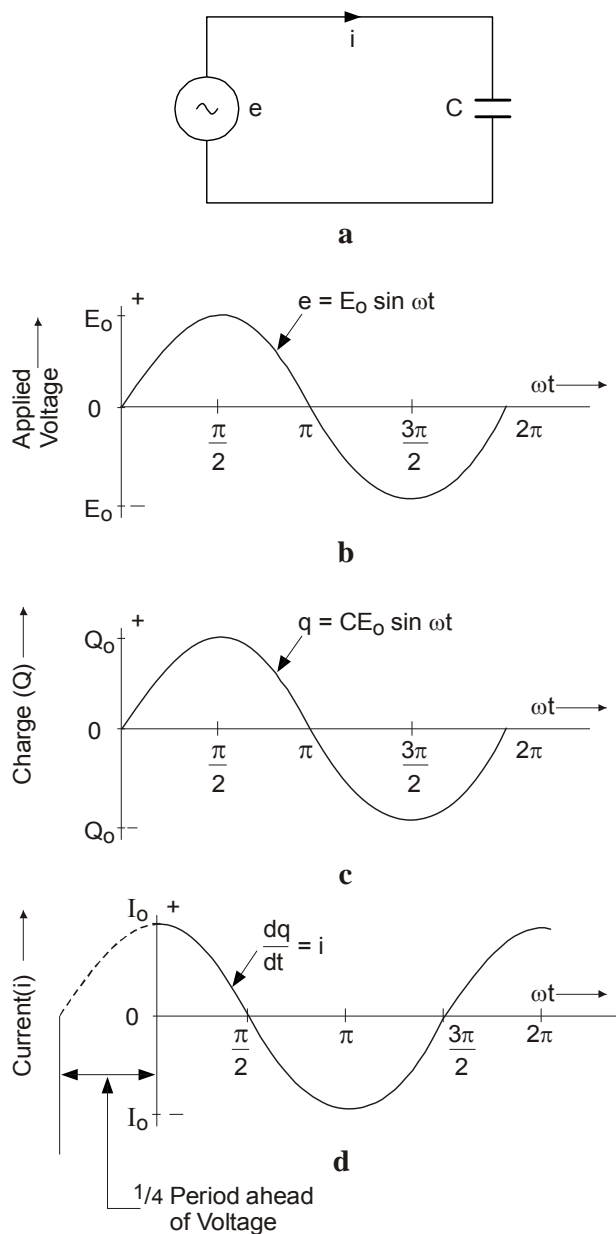
42. In a pure capacitance the ratio of voltage to current is:

$$\frac{E_0}{I_0} = \frac{E}{I} = \frac{1}{\omega C} = \frac{1}{2\pi fC} \text{ or, transposing}$$

$$I = E2\pi fC$$

That is, in a circuit having capacitance only, the current is directly proportional to the applied voltage, the frequency, and the capacitance.

**14-7 Fig 24 Phase Relationship in a Capacitance**



43. **Capacitive Reactance.** It will be recalled, from Volume 14, Chapter 4, that capacitance opposes any change in the value of voltage applied to a capacitor, thus when the applied voltage is alternating the capacitor presents an opposition at all times. This opposition is called capacitive reactance ( $X_C$ ), and, as with inductive reactance, the term reactance is used rather than resistance because the opposition to the change depends upon frequency.

44. Capacitive reactance is expressed in ohms, and is given by:

$$X_C = \frac{E}{I} = \frac{1}{2\pi fC}$$

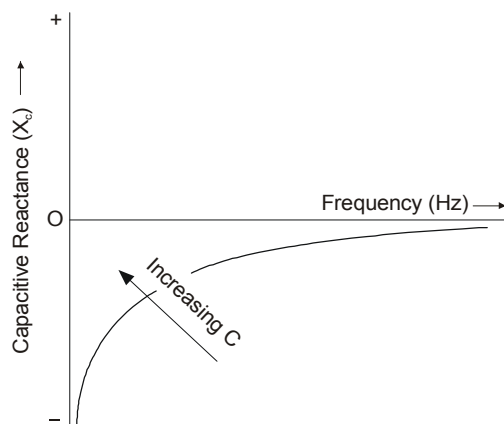
Thus, capacitive reactance is inversely proportional to frequency. A graph of capacitive reactance against frequency is shown at Fig 25. The reactance of a capacitance is assumed negative (opposite to that of an inductance), and is normally given as  $X_C = 1/\omega C$ ; it decreases as the capacitance increases, and as the frequency increases. Thus, for a  $2\mu\text{F}$  capacitor:

$$\text{when } f \text{ is } 50 \text{ Hz, } X_C = \frac{1}{2\pi fC} = \frac{10^6}{2\pi \times 50 \times 2} = 1,600 \, \Omega$$

$$\text{when } f \text{ is } 1 \text{ kHz, } X_C = \frac{10^6}{2\pi \times 10^3 \times 2} = 80 \, \Omega$$

$$\text{when } f \text{ is } 1 \text{ MHz, } X_C = \frac{10^6}{2\pi \times 10^6 \times 2} = .08 \, \Omega$$

14-7 Fig 25 Variation of  $X_C$  with Frequency

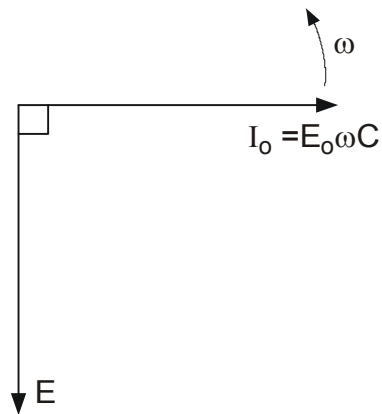


45. **Power.** The power curve for a purely capacitive circuit is derived as usual by multiplying together corresponding values of current and voltage to find the instantaneous power, and plotting the products. The power curve is similar to that derived in Fig 22 for an inductive circuit, but, as in this case current leads voltage, inverted. Again the total power is zero.

46. **Conclusions.** In a purely capacitive circuit:

- a. The current leads the voltage by  $90^\circ$  ( $\pi/2$  radians).
- b. The ratio of voltage to current gives the capacitive reactance  $X_C$ .

These two statements can be shown by drawing a vector diagram as at Fig 26.

**14-7 Fig 26 Phases of Current and Voltage in a Pure Capacitance****Summary**

47. The relative phase of current and voltage in a capacitor and in an inductor can be remembered from the word 'CIVIL', where C is the capacitance, I the current, V the voltage, and L the inductance. CIVIL then indicates that in a capacitor the current leads the voltage; in an inductor, the current lags the voltage.

**CHAPTER 8 - AC CIRCUITS**

<b>Contents</b>	<b>Page</b>
Introduction .....	2
SERIES CIRCUITS .....	2
Resistance and Inductance in Series .....	2
Resistance and Capacitance in Series .....	4
Inductance and Capacitance in Series .....	6
Resistance, Inductance and Capacitance in Series .....	7
Power .....	10
Series Circuit Resonance .....	14
Voltage Magnification .....	16
PARALLEL CIRCUITS .....	17
Introduction .....	17
Resistance and Inductance in Parallel .....	17
Resistance and Capacitance in Parallel .....	18
Inductance and Capacitance in Parallel .....	19
Tuned Circuit with Resistance .....	19
Current Magnification .....	22
<b>Table of Figures</b>	
14-8 Fig 1 R and L in Series .....	2
14-8 Fig 2 Phase Relationships, R and L in Series .....	3
14-8 Fig 3 Magnitude and Phase of V .....	3
14-8 Fig 4 Impedance Triangle, R and L in Series .....	4
14-8 Fig 5 R and C in Series .....	5
14-8 Fig 6 Phase Relationship, R and C in Series .....	5
14-8 Fig 7 Impedance Triangle, R and C in Series .....	6
14-8 Fig 8 L and C in Series .....	7
14-8 Fig 9 R, L and C in Series .....	7
14-8 Fig 10 Resultant Applied Voltage, R, L and C in Series .....	8
14-8 Fig 11 Impedance Triangle, R, L and C in Series .....	9
14-8 Fig 12 Effect of Variation of Frequency .....	9
14-8 Fig 13 Working and Wattless Components of Current .....	11
14-8 Fig 14 Power Factor of a Coil .....	13
14-8 Fig 15 Power Factor of a Capacitor .....	13
14-8 Fig 16 Power Factor of a Capacitor .....	14
14-8 Fig 17 Series Resonance .....	14
14-8 Fig 18 Reactance Sketch for a Series Tuned Circuit .....	15
14-8 Fig 19 Variation of Z with Frequency .....	15
14-8 Fig 20 Voltage Magnification .....	17
14-8 Fig 21 R and L in Parallel .....	18
14-8 Fig 22 R and C in Parallel .....	19
14-8 Fig 23 L and C in Parallel .....	19
14-8 Fig 24 Practical Parallel Tuned Circuit .....	20
14-8 Fig 25 Vector Relationships in a Parallel Tuned Circuit .....	20
14-8 Fig 26 Effect of Variation of Frequency .....	21
14-8 Fig 27 Effect of Varying $X_C$ and $X_L$ .....	22
14-8 Fig 28 Supply and Circulating Current .....	22
14-8 Fig 29 Current Magnification .....	23

## Introduction

1. In the last chapter ideal components in AC circuits were examined. In this chapter components in combination as they actually occur in circuits are discussed.
2. Circuits containing combinations of resistance, inductance and capacitance in series and in parallel are considered in detail, and the phase and magnitude relationships of circuit voltages and currents are illustrated for various typical circuits.

### Notes:

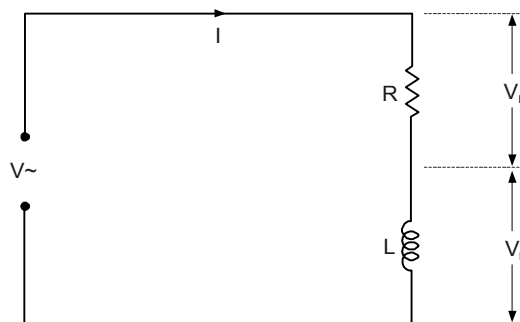
1. Unless otherwise specified it should be assumed that rms values of current and voltage are intended.
2. As this is a general theoretical treatment, units are not always specified. It should be assumed that quantities are in compatible units, viz current in amperes, emf and PD in volts, frequency in hertz, resistance in ohms, inductance in henrys, and capacitance in farads.

## SERIES CIRCUITS

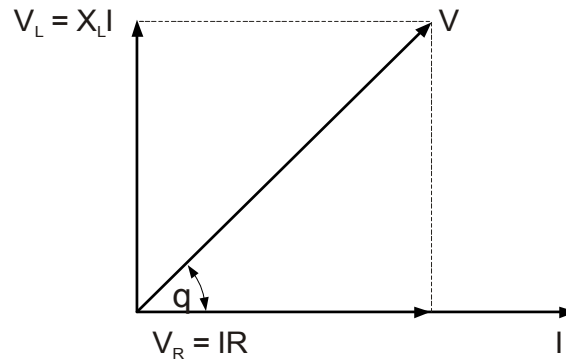
### Resistance and Inductance in Series

3. In Fig 1 a sinusoidal alternating voltage of rms value  $V$  volts and frequency  $f$  hertz is applied across a circuit consisting of resistance  $R$  ohms and inductance  $L$  henrys connected in series. A corresponding rms current  $I$  amperes is established in the circuit. With components in series, the current is the same at all points in the circuit, and, for this reason, the current vector is taken as the reference when constructing a vector diagram.

14-8 Fig 1 R and L in Series



4. The vector diagram for this circuit is shown in Fig 2 and is constructed in the following manner:
  - a. The current vector  $I$  is drawn to form the reference line.
  - b. The voltage drop across  $R$  is the product of the current and the resistance. It is in phase with the current and has an rms value  $V_R = IR$ . This vector is drawn to any convenient scale in line with the current vector.
  - c. The voltage across  $L$  is the product of the current and the inductive reactance  $X_L$ . It leads the current by  $90^\circ$  and has an rms value  $V_L = IX_L$ . This vector is drawn to the same scale as  $V_R$ , leading the current by  $90^\circ$ .
  - d. The applied voltage  $V$  is found by applying the analogous parallelogram of forces rule. This voltage is seen to lead the current by an angle  $\theta$  (the phase angle).

**14-8 Fig 2 Phase Relationships, R and L in Series**

5. **Application of Pythagoras' Theorem.** The magnitude and phase of the applied voltage  $V$  is found by considering the triangle ABC in Fig 3. This is a right-angled triangle to which Pythagoras' theorem applies. Thus, from Fig 3:

$$V^2 = V_R^2 + V_L^2, \text{ and } V = \sqrt{V_R^2 + V_L^2}$$

Alternatively, substituting for  $V_R$  and  $V_L$ ,

$$V^2 = I^2 R^2 + I^2 X_L^2$$

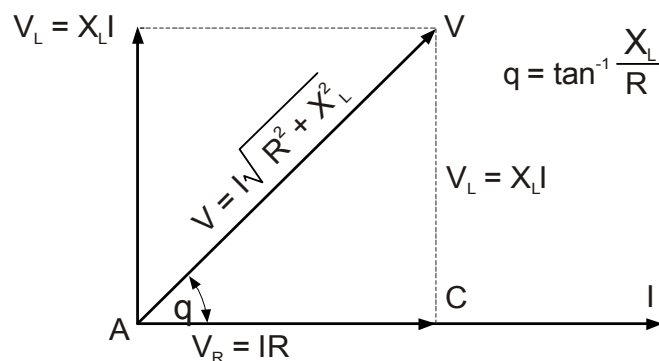
$$\therefore V = I \sqrt{R^2 + X_L^2}$$

Also,  $V$  leads  $I$  by an angle  $\theta$  where:

$$\tan \theta = \frac{V_L}{V_R} = \frac{X_L}{R}$$

$$\therefore \theta = \tan^{-1} \frac{\text{Reactance}}{\text{Resistance}}$$

(Where  $\tan^{-1}$  means "the angle whose tangent is given by".)

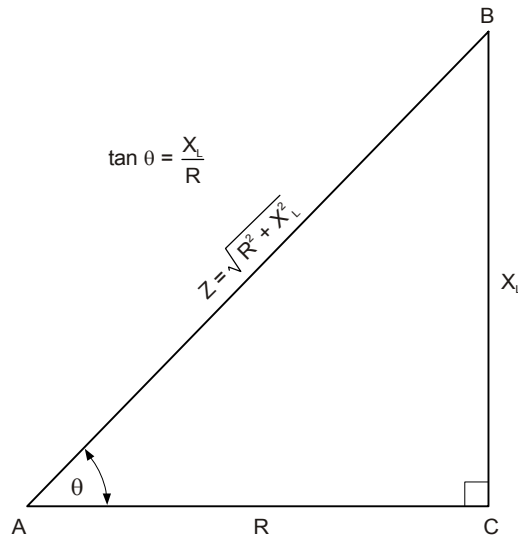
**14-8 Fig 3 Magnitude and Phase of  $V$** 

6. **Impedance.** The total opposition to current in a circuit containing resistance and reactance in combination is termed impedance (symbol  $Z$ ) and is measured in ohms. For the whole circuit  $I = V/Z$ . Using  $Z = V/I$ , which applies to any circuit, for a series circuit containing resistance and inductance where the voltage  $V$  is  $I \sqrt{R^2 + X_L^2}$  the impedance is:

$$Z = \frac{V}{I} = \sqrt{R^2 + X_L^2}$$

7. **The Impedance Triangle.** Although  $Z$  is composed of both resistance and reactance, it is not merely the sum of these, as the phases of current and voltage in each of them are different. An impedance triangle can be constructed from the vector diagram of Fig 2 by dividing each voltage vector by  $I$ . The result is shown in Fig 4, where the base represents the resistance, the perpendicular the reactance, and the hypotenuse the impedance. The tan of the phase angle  $Z$  is given by  $X_L/R$ .

**14-8 Fig 4 Impedance Triangle, R and L in Series**



8. **Summary.** In a circuit containing resistance and inductance in series:

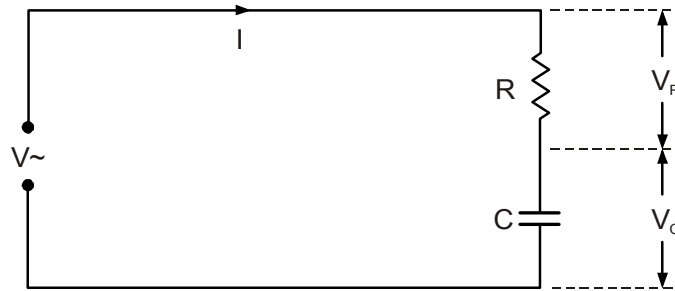
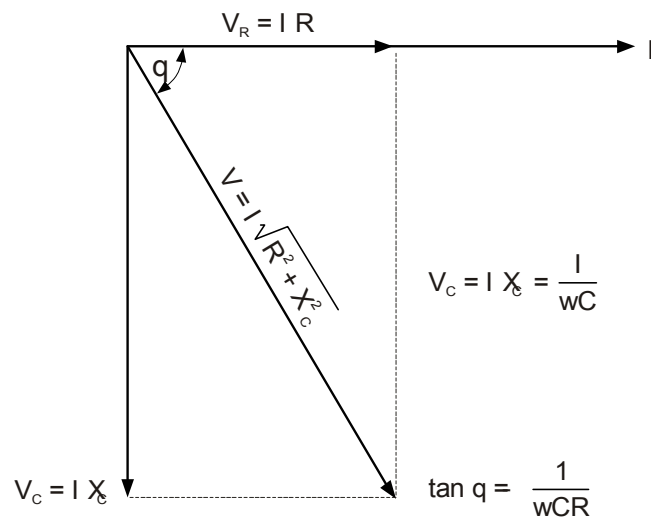
- The applied voltage  $V = \sqrt{V_R^2 + V_L^2}$  volts.
- The magnitude of the current  $I = \frac{V}{Z} = \frac{V}{\sqrt{R^2 + X_L^2}}$  amperes.
- The phase of the applied voltage relative to the current is found from  $\tan \theta = X_L/R$ ,  $\tan \theta$  being positive indicating that  $V$  leads  $I$ .
- The impedance of the circuit is  $Z = \sqrt{R^2 + X_L^2}$  ohms, where  $X_L$  (the inductive reactance), equals  $2\pi fL$  (i.e.  $\omega L$ ).

### Resistance and Capacitance in Series

9. With a current of rms value  $I$  amperes established in the circuit of Fig 5, components of the applied voltage  $V$  appear across  $R$  and across  $C$ . The vector diagram (Fig 6) is constructed as follows:

- The current  $I$  is common to  $R$  and  $C$  and the vector  $I$  is the reference vector.
- The voltage across  $R$  is  $V_R = IR$ , and is in phase with  $I$ .
- The voltage across  $C$  is  $V_C = IX_C = -I/\omega C$ , and it lags the current by  $90^\circ$ . Its magnitude decreases as the frequency increases.
- The resultant applied voltage  $V$  is found by applying the analogous parallelogram of forces rule. This voltage is then seen to lag the current by an angle  $\theta$ .



**14-8 Fig 5 R and C in Series****14-8 Fig 6 Phase Relationship, R and C in Series**

10. **Application of Pythagoras' Theorem.** Applying Pythagoras' theorem to Fig 6:

$$V^2 = V_R^2 + V_C^2, \text{ and } V = \sqrt{V_R^2 + V_C^2}$$

Or, substituting:

$$V^2 = I^2 R^2 + I^2 X_C^2, \text{ and } V = I \sqrt{R^2 + X_C^2}.$$

Also,  $V$  lags  $I$  by an angle  $\theta$  where:

$$\tan \theta = \frac{V_C}{V_R} = \frac{V_C}{IR} = \left( -\frac{1}{\omega CR} \right)$$

$$\therefore \theta = \tan^{-1} \left( -\frac{\text{reactance}}{\text{resistance}} \right)$$

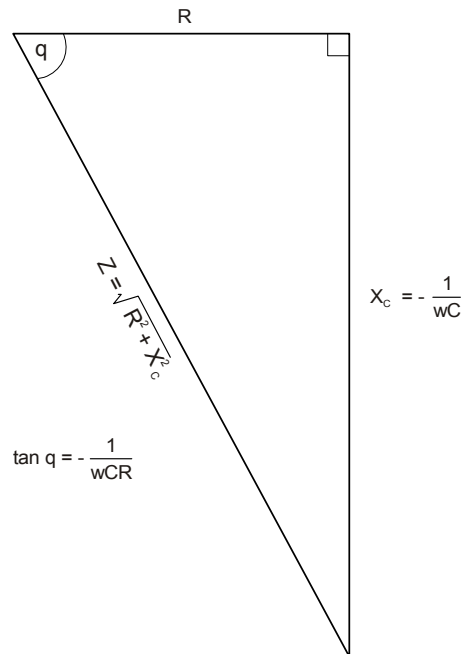
Note that  $\tan \theta$  is negative in a capacitive circuit, indicating that  $V$  lags  $I$ .

11. **Impedance.** The impedance of this circuit is:

$$Z = \frac{V}{I} = \sqrt{R^2 + X_C^2}$$

The corresponding impedance triangle is shown in Fig 7.

**14-8 Fig 7 Impedance Triangle, R and C in Series**



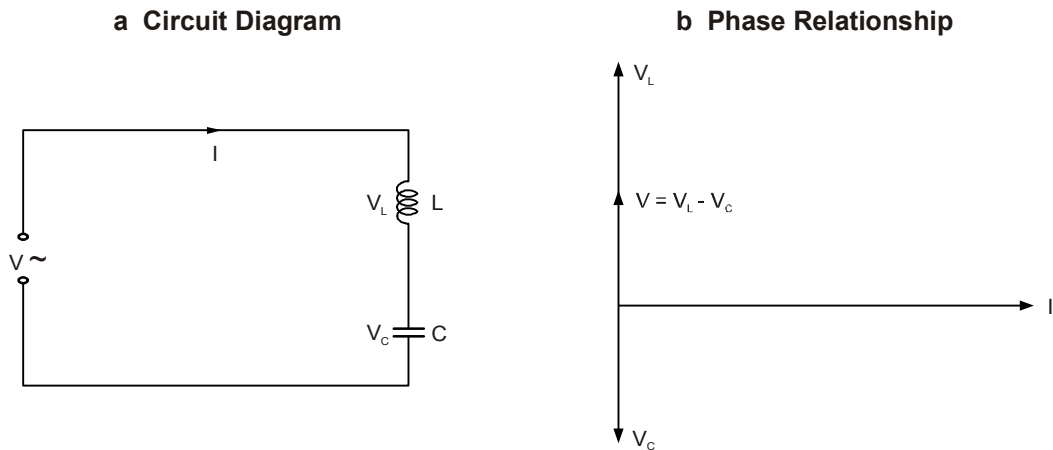
12. **Summary.** In a circuit containing resistance and capacitance in series:

- The applied voltage  $V = \sqrt{V_R^2 + V_C^2}$  volts.
- The magnitude of the current is  $I = \frac{V}{Z} = \frac{V}{\sqrt{R^2 + X_C^2}}$  amperes.
- The phase of the current relative to the applied voltage is found from  $\tan \theta = X_C/R$ . The current leads the voltage in a capacitive circuit.
- The impedance of the circuit is:  $Z = \sqrt{R^2 + X_C^2}$  ohms, where  $X_C$ , the capacitive reactance, equals  $1/2\pi fC$ .

### Inductance and Capacitance in Series

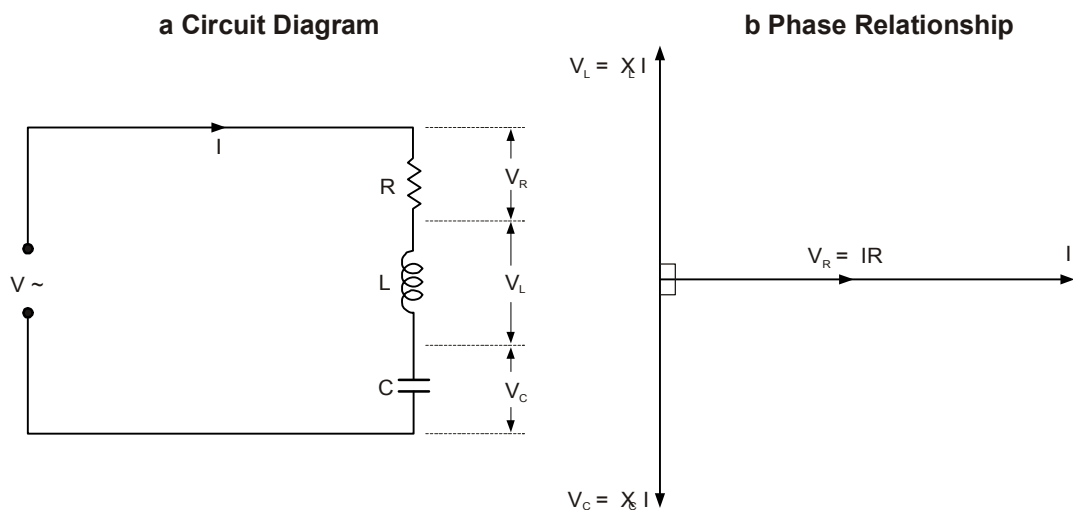
13. In an AC circuit containing only L and C in series (Fig 8a), the voltage  $V_L$  across the inductance leads the current by  $90^\circ$ , and voltage  $V_C$  across the capacitance lags the current by  $90^\circ$ . Thus  $V_L$  and  $V_C$  are  $180^\circ$  out of phase as shown in Fig 8b, and oppose each other so that the total applied voltage  $V$  is their difference. In Fig 8  $V_L$  is taken to be larger than  $V_C$  and  $V$  is leading  $I$  by  $90^\circ$ . However, if the frequency of the applied voltage and values of L and C are such that  $X_C > X_L$ , then  $V_C > V_L$  and  $V$  lags  $I$  by  $90^\circ$ .

14. Ohm's law applies to each part of the circuit and to the whole circuit:  $V_L = IX_L$ ;  $V_C = IX_C$ ;  $V = IZ$ . Since  $V_L$  and  $V_C$  oppose each other,  $X_L$  and  $X_C$  act in opposite directions and the impedance of the circuit is their difference. Thus  $Z = X_L - X_C$ , or  $X_C - X_L$ , depending on which is the larger. If  $X_L > X_C$ , then  $Z$  is an inductive reactance. If  $X_C > X_L$ , then  $Z$  is a capacitive reactance.

**14-8 Fig 8 L and C in Series****Resistance, Inductance and Capacitance in Series**

15. Fig 9a shows a circuit consisting of  $R$  ohms,  $L$  henrys and  $C$  farads connected in series across an alternating supply of frequency  $f$  Hz. With a current of rms value  $I$  amperes established in the circuit, voltages (in volts) across each component are as follows:

- Across  $R$ ,  $V_R = IR$  and is in phase with  $I$ .
- Across  $L$ ,  $V_L = IX_L$  and leads  $I$  by  $90^\circ$ . Its magnitude increases with frequency.
- Across  $C$ ,  $V_C = IX_C$  and lags  $I$  by  $90^\circ$ . Its magnitude decreases as the frequency increases.

**14-8 Fig 9 R, L and C in Series**

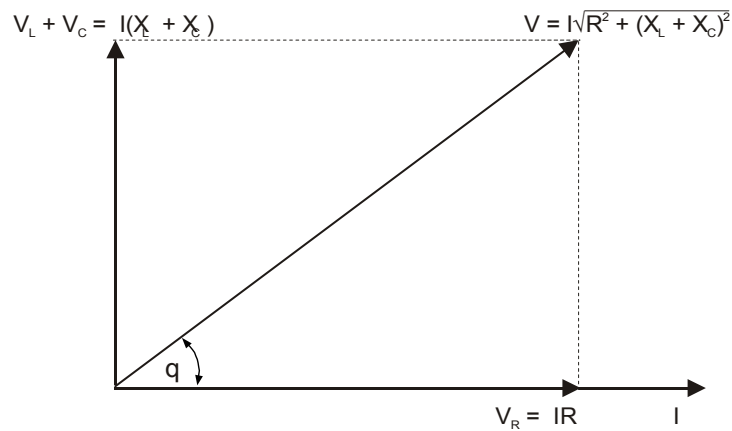
16. The resultant vector diagram is constructed as shown in Fig 9b, the current  $I$ , which is common to all components, being the reference vector. By Kirchoff's second law the vectorial resultant of  $V_R$ ,  $V_L$ , and  $V_C$  must equal the applied voltage  $V$ . To obtain this:

- Add the two vectors  $V_L = IX_L$  and  $V_C = IX_C$  to give a single resultant vector  $I(X_C + X_L)$ . Since the capacitive reactance is negative with respect to the inductive reactance and is given by:

$$X_C = -\frac{1}{\omega C}, \text{ the resultant vector is } I\left(\omega L - \frac{1}{\omega C}\right).$$

- b. Apply the rule of the parallelogram of forces to vector  $V_R$  and the combined vector  $(V_L + V_C)$  to obtain the resultant applied voltage vector  $V$ . This is shown in Fig 10.

**14-8 Fig 10 Resultant Applied Voltage, R, L and C in Series**



17. **Application of Pythagoras' Theorem.** Apply Pythagoras' theorem to Fig 10:

$$V^2 = V_R^2 + (V_L + V_C)^2 \text{ and } V = \sqrt{V_R^2 + (V_L + V_C)^2}$$

Or, substituting:

$$V^2 = I^2 R^2 + (IX_L + IX_C)^2 \text{ and } V = I\sqrt{R^2 + (X_L + X_C)^2}$$

$$\therefore V = I\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}$$

In this case,  $V$  leads  $I$  by an angle  $\theta$  where:

$$\tan \theta = \frac{V_L + V_C}{V_R} = \frac{(X_L + X_C)}{R}$$

$$\therefore \tan \theta = \frac{(\omega L - 1/\omega C)}{R}$$

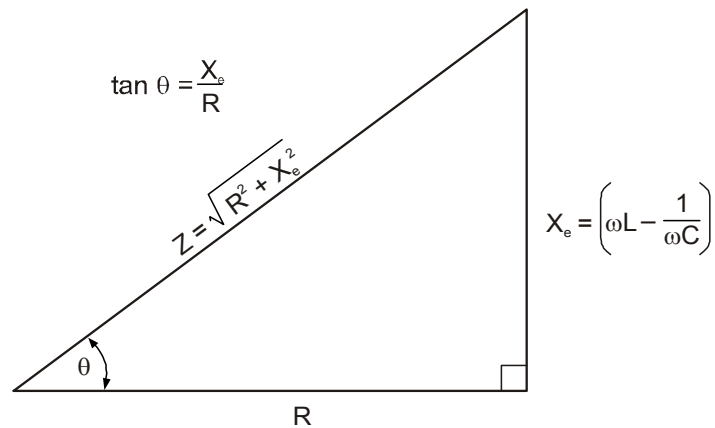
$$\theta = \tan^{-1} \left( \frac{\text{effective reactance}}{\text{resistance}} \right)$$

The factor  $\tan \theta$  is positive in this instance since  $X_L$  is greater than  $X_C$ .

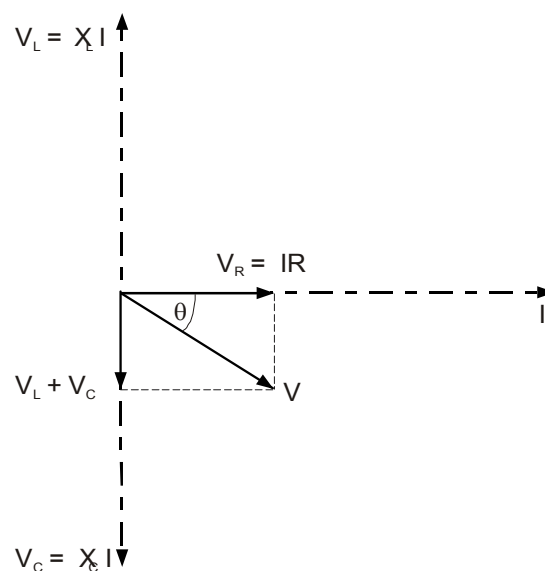
18. **Impedance.** The impedance of this circuit is:

$$Z = \frac{V}{I} = \sqrt{R^2 + (X_L + X_C)^2}$$

The corresponding impedance triangle is shown in Fig 11. It should be noted that  $X_L$  is a positive reactance and  $X_C$  is a negative reactance so that it is, in fact, their numerical difference which gives the magnitude of the effective reactance ( $X_e$ ) in the circuit.

**14-8 Fig 11 Impedance Triangle, R, L and C in Series**

19. **Variation of Frequency.** A variation in the frequency of the supply has an effect on the results obtained. The vector diagrams of Figs 9b and 10 apply when the frequency of the supply voltage is such that  $X_L$  is greater than  $X_C$ . As the frequency is reduced,  $X_L$  is reduced and  $X_C$  is increased. Thus the vector  $V_L = IX_L$  is correspondingly reduced in amplitude, and the vector  $V_C = IX_C$  increased. The condition where the frequency of the supply voltage is such that  $X_C$  is greater than  $X_L$  is shown in Fig 12, in which it is seen that the resultant voltage  $V$  now lags the current  $I$ . Thus the applied voltage  $V$  either leads or lags the current  $I$  depending on whether  $V_L$  is greater than  $V_C$  (ie  $X_L > X_C$ ), or  $V_C$  is greater than  $V_L$  (ie  $X_C > X_L$ ). If the voltage leads the current ( $\tan \theta$  positive) the circuit is 'inductive'; if the voltage lags the current ( $\tan \theta$  negative) the circuit is 'capacitive'.

**14-8 Fig 12 Effect of Variation of Frequency**

20. **Variation of L or C.** Similar results to those in the preceding paragraphs are obtained if the frequency of the supply is fixed and the value of  $L$  or  $C$  is altered. With a constant value of  $\omega$  (ie of  $2\pi f$ ) the inductive reactance  $\omega L$  can be altered by varying the value of  $L$ . Similarly, a variation in  $C$  will alter the capacitive reactance ( $1/\omega C$ .) In this way,  $X_L$  or  $X_C$  can be varied to give resultant vector diagrams similar to Figs 10 and 11.

21. **Summary.** In a circuit containing resistance, inductance and capacitance in series:

- a. The magnitude of the applied voltage is  $V = \sqrt{V_R^2 + (V_L \sim V_C)^2}$ , noting that  $V_C$  is negative.
- b. The magnitude of the current is  $I = V/Z = V/\sqrt{R^2 + (X_L \sim X_C)^2}$ .
- c. The phase of the applied voltage relative to the current is found from  $\tan \theta = (\omega L - 1/\omega C)/R$ , or  $(X_L \sim X_C)/R$  noting that  $X_C$  is negative, and the angle is either positive or negative depending on the relative values of  $X_L$  and  $X_C$ .
- d. The impedance of the circuit is  $Z = \sqrt{R^2 + (X_L \sim X_C)^2}$ , noting that  $X_C$  is negative.

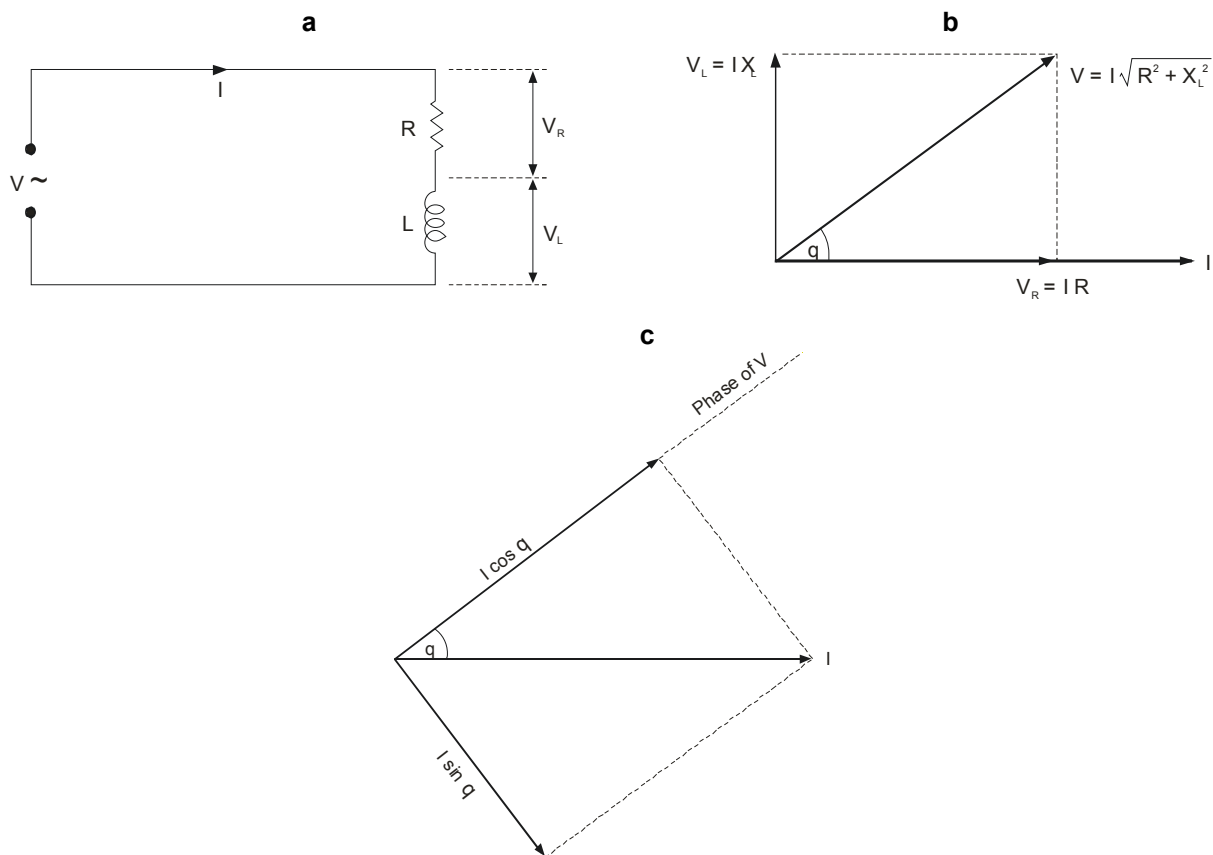
### Power

22. When an alternating voltage is applied across a pure resistance  $R$ , current and voltage are always in phase. The power developed in  $R$  is  $VI$  watts, where  $V$  and  $I$  are rms values.

23. In a coil which is considered to be a pure inductance, there is no dissipation of energy as heat when the coil is connected across an alternating supply of voltage. The current and voltage are  $90^\circ$  out of phase, and the energy supplied ( $LI^2/2$ ) during that part of the cycle when the current is growing to its peak value is stored in the magnetic field, and restored to the source as the current decays to zero. Similarly, in a pure capacitance the current and voltage are  $90^\circ$  out of phase, and the energy supplied ( $CV^2/2$ ) to the capacitance during that part of the cycle when the voltage is building up to its peak value is stored in the electric field, and restored to the source as the voltage falls to zero. Thus, when an alternating voltage is applied across a pure reactance, either inductive or capacitive, no power is developed at all and the current is said to be 'wattless'.

24. In an actual circuit containing both resistance and reactance, there is a phase difference ( $\theta$ ) between the current and the voltage. The vector diagram for the circuit of Fig 13a is given in Fig 13b, from which it is seen that the current  $I$  lags the applied voltage ( $V$ ) by an angle  $\theta$ . It was shown in Volume 14, Chapter 7 that a vector  $I$ , at an angle  $\theta$  to the reference, can be resolved into two components at right angles to each other in terms of  $I$  and  $\theta$ . This is shown in Fig 13c. The component of current  $I$  which is in phase with the applied voltage (termed the in-phase component of  $I$ ) is  $I \cos \theta$ ; the out-of-phase component of  $I$  at right angles to the applied voltage is  $I \sin \theta$ . The former current develops a power of  $VI \cos \theta$  watts. The latter develops no power at all since it is  $90^\circ$  out of phase with the applied voltage; for this reason it is termed the wattless component of current.

14-8 Fig 13 Working and Wattless Components of Current



25. **Power Factor.** The factor  $\cos \theta$  is known as the 'power factor' of a circuit. The product  $VI$  is that of the rms values of voltage and current and is termed the voltamperes (VA) or the apparent power in a circuit. The true power in a circuit, where  $V$  and  $I$  are out of phase, is given by the product of the applied voltage and the in-phase component of current. Thus:

$$P = VI \cos \theta$$

$$\cos \theta = \frac{P}{VI}$$

$$\therefore \text{power factor} = \frac{\text{true power}}{\text{apparent power}}$$

The power factor may be found by measuring the true power with a wattmeter, and the apparent power with a voltmeter and ammeter. Referring again to Fig 13b it is seen that:

$$\text{voltage } V = I\sqrt{R^2 + X_L^2} = IZ.$$

$$\therefore \cos \theta = \frac{IR}{IZ} = \frac{R}{Z} = \text{power factor}$$

This gives an alternative expression for the power factor in terms of the components in a circuit. Summarizing, for a series AC circuit:

- True power = power actually consumed in the resistance of a circuit =  $VI \cos \theta$ .
- Apparent power = volts  $\times$  amperes =  $VI$ .

c. Power factor =  $\frac{\text{true power}}{\text{apparent power}} = \cos \theta = \frac{R}{Z}$ .

d. When V and I are in phase,  $\theta = 0$ ,  $\cos \theta = 1$ , and true power = apparent power.

e. When V and I are  $90^\circ$  out of phase,  $\theta = 90^\circ$ ,  $\cos \theta = 0$ , and true power = zero.

26. **Adjustment of Power Factor.** In a pure resistance, the power factor  $\cos \theta = R/Z = 1$ . In a pure reactance, it is zero. In a practical circuit, it lies between these extremes. The value of power factor required by a circuit depends on the purpose of the circuit. Where the power developed must be kept to a minimum, the power factor should be as nearly zero as possible (ie R should be small). Where a large power is required to be developed, eg electric fires, the power factor should be as near to unity as possible (i.e. X should be small).

27. **Power Factor of One.** Where a power factor of unity is required, the out-of-phase component of current must be zero (ie the reactive component must be zero). When the circuit is inductive, the current will lag the voltage by an angle  $\theta$ , and the power factor will be less than unity. This can be adjusted by inserting a capacitive component of such value that it balances the inductive component. The current and voltage will then be in phase and the power factor will be unity.

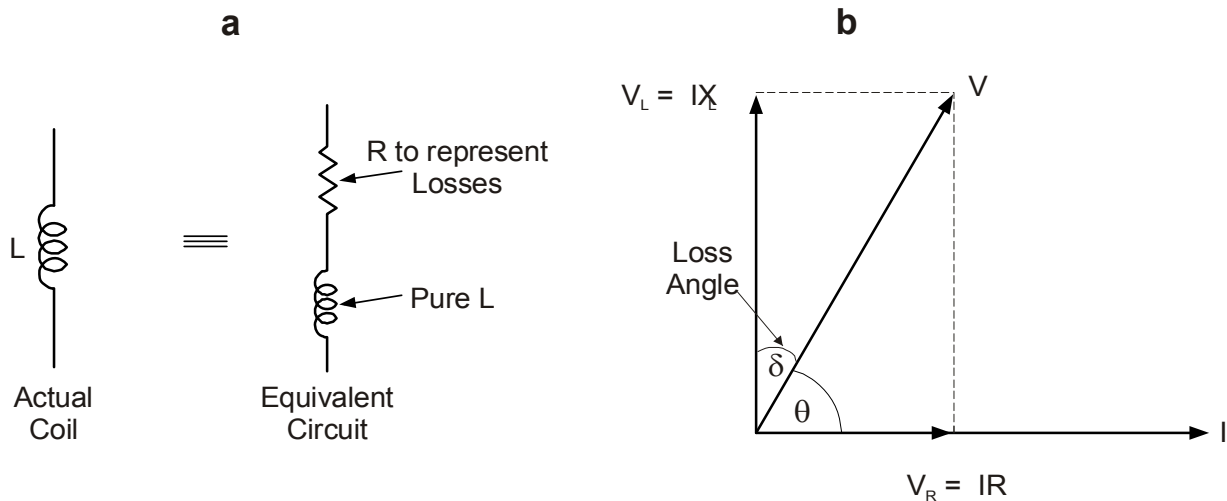
28. **Power Losses in Components.** The power factor of an inductor or a capacitor should, in theory, be zero. That is, these components should be pure reactances. This cannot apply in practice however, since power losses are developed in the components. Most of the losses listed below increase with the frequency of the supply. They occur mainly in high frequency circuits, and because of this, are more important in radio engineering than in power engineering.

29. **Losses in Inductors.** Power losses occur in inductors because of:

- a. Eddy current in the cores.
- b. Hysteresis in the cores.
- c. Skin effect in the coils.
- d. Proximity effect of conductors in the near vicinity.
- e. Ohmic resistance of the windings.

All of these losses can be grouped together by supposing that the inductor has in series with it a resistance R ohms, such that the power loss is  $I^2 R$  watts, where I amperes is the rms value of the current. An inductor can thus be represented as L henrys and R ohms in series, where L is a pure inductance and R represents the losses (Fig 14a). The vector diagram is then as shown in Fig 14b, from which it is seen that I no longer lags V by  $90^\circ$  but by an angle  $\theta$ , this being less than  $90^\circ$  by  $\delta$  (the loss angle). The inductor has, therefore, a power factor of  $\cos \theta$ .

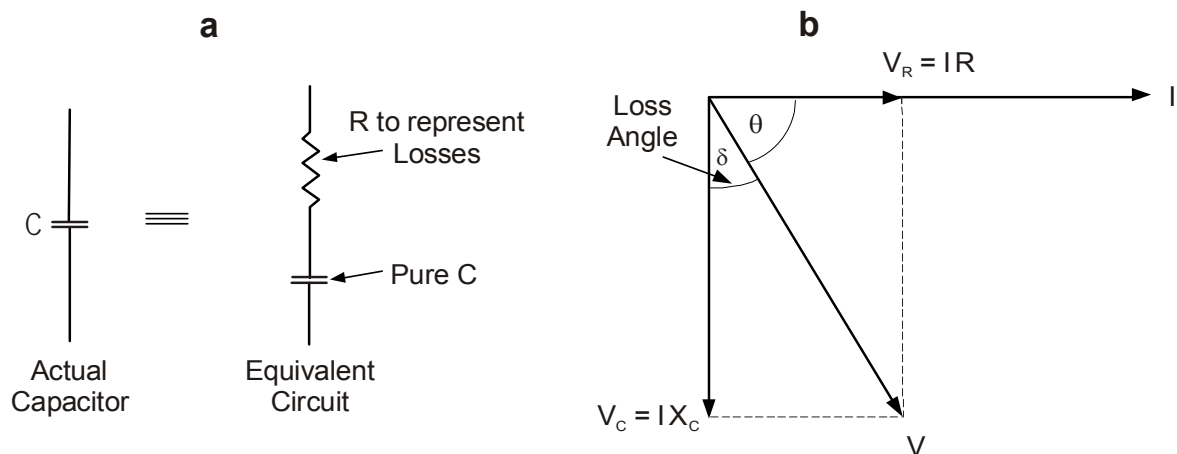


**14-8 Fig 14 Power Factor of a Coil**

30. **Losses in Capacitors.** Power losses occur in capacitors because of:

- Leakage through the dielectric.
- Dielectric hysteresis or absorption.
- Brush discharge.
- Ohmic resistance of the connecting leads.
- Skin effect in the plates and connecting leads.

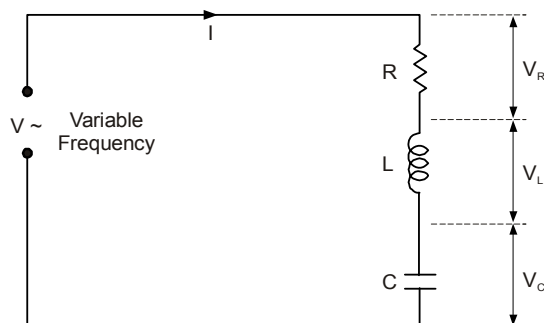
Remarks similar to those for the inductor apply to the capacitor, and a capacitor can be represented as C and R in series, where C is a pure capacitance and R represents the losses (Fig 15a). The vector diagram is then as shown in Fig 15b, from which it is seen that I leads V by an angle  $\theta$ , this being less than  $90^\circ$  by the angle  $\delta$  (the loss angle). A capacitor has, therefore, a power factor  $\cos \theta$ . In practice it is nearer zero than that of an inductor and in normal circuits of capacitor and inductor in series, most of the losses are in the inductor.

**14-8 Fig 15 Power Factor of a Capacitor**

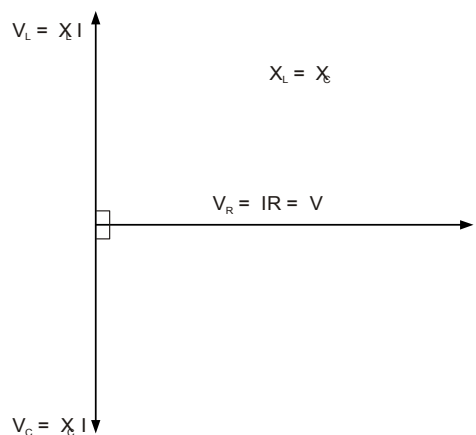
### Series Circuit Resonance

31. Consider the circuit shown in Fig 16 where R is the combined loss resistance of L and C. When the frequency of the supply is such that the capacitive reactance  $X_C$  is greater than the inductive reactance  $X_L$ , the impedance is capacitive and the phase angle negative indicating that the applied voltage is lagging the current. When the frequency of the supply is such that  $X_L$  is greater than  $X_C$ , the impedance is inductive and the phase angle positive, indicating that the applied voltage is leading the current. In between these conditions, at some particular frequency,  $X_L$  will equal  $X_C$ , and  $V_L$  will equal  $V_C$ . This is shown vectorially in Fig 17. The circuit is then said to be at resonance. Where the frequency of the supply is fixed, either L or C can be varied to give the condition  $X_L = X_C$ . The circuit has then been tuned to resonance with the supply frequency.

**14-8 Fig 16 Power Factor of a Capacitor**



**14-8 Fig 17 Series Resonance**

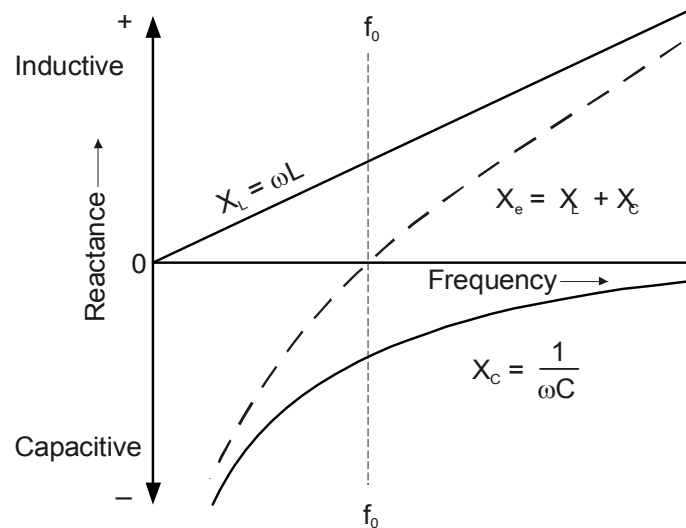


32. **Conditions at Resonance.** At resonance:

- $\omega L = \frac{1}{\omega C}$ , and the impedance  $Z$  is given by:  $Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} = R$  ohms. Thus, at resonance in a series tuned circuit,  $Z$  is a minimum and equal to  $R$  ohms. In a well-designed circuit,  $R$  will be small, being only the loss resistance of  $L$  and  $C$ .
- The current  $I = \frac{V}{Z} = \frac{V}{R}$  amperes has a maximum value.
- The phase angle  $\theta$ , given by  $\tan \theta = \frac{(\omega L - 1/\omega C)}{R}$ , is zero. Thus, at resonance in a series circuit, the current and the applied voltage are in phase.

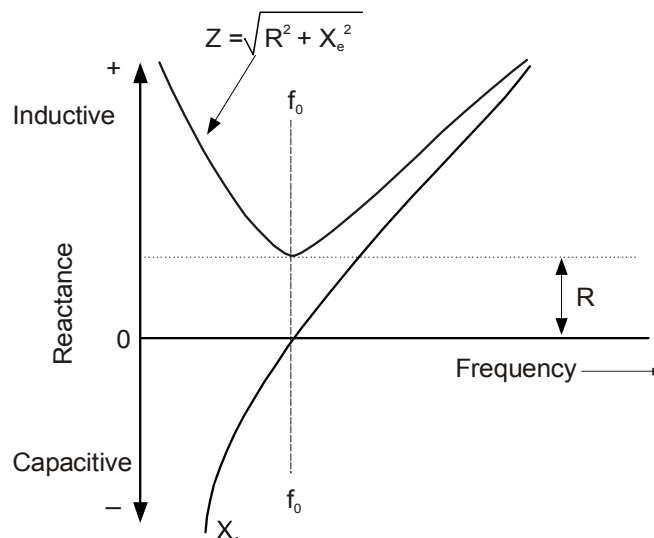
33. **Reactance Sketches.** A reactance sketch is a graph relating reactance and frequency. The reactance sketches for an inductance and for a capacitance are shown in Volume 14, Chapter 7. In a series circuit where both capacitance and inductance are present, the reactance sketches can be combined to give the total variation of reactance with frequency. This is shown in Fig 18. From this graph it is seen that at the resonant frequency ( $f_0$ ),  $X_L = X_C$  and the total reactance is zero. The circuit is then purely resistive with the current and the applied voltage in phase.

14-8 Fig 18 Reactance Sketch for a Series Tuned Circuit



34. **Variation of Impedance with Frequency.** If the graph for resistance  $R$  (which for practical purposes remains constant with frequency) is combined with the graph for the total effective reactance  $X_e$ , a curve for impedance  $Z$  against frequency can be obtained. To plot this graph the expression  $Z = \sqrt{R^2 + X_e^2}$  ohms is used and as shown in Fig 19,  $Z$  is all above the horizontal axis. At the resonant frequency  $f_0$  the impedance  $Z$  is minimum and equal to  $R$  ohms. Either side of resonance the impedance rises, and, as can be seen from the reactance curve, below resonance the circuit has capacitive reactance and above resonance the circuit has inductive reactance.

14-8 Fig 19 Variation of  $Z$  with Frequency



35. **Resonant Frequency.** At resonance,  $X_L = X_C$ , i.e.

$$2\pi f_0 L = \frac{1}{2\pi f_0 C} \quad \therefore f_0^2 = \frac{1}{4\pi^2 LC} \quad \therefore f_0 = \frac{1}{2\pi\sqrt{LC}}$$

where  $f_0$  = resonant frequency in hertz,  $L$  = inductance in henrys,  $C$  = capacitance in farads

This gives the resonant frequency of a series tuned circuit. By altering either  $L$  or  $C$  the frequency at which resonance occurs is changed. It should be noted that  $R$  has no effect on this.

### Voltage Magnification

36. **Selectivity.** Selectivity is the property of a tuned circuit which makes it responsive to a particular frequency. Selectivity is dealt with more fully in the Annex, but it is sufficient here to note that it is represented by the symbol  $Q_0$ , and defined as the ratio of reactance to circuit resistance at the resonant frequency, i.e.  $Q_0 = 2\pi f_0 L/R$  (or  $\omega L/R$ ).

37. At resonance, in a series RLC circuit, the impedance is a minimum and equal to the resistance  $R$ . The current  $I$  is a maximum and equal to  $V/R$ . Voltages across the components shown in Fig 20a are as follows:

a. **Resistor.** The voltage across the resistance at resonance is:

$$V_R = IR = \frac{V}{R} \times R = V \text{ volts}$$

$$\therefore V_R = \text{applied voltage}$$

b. **Inductor.** The voltage across the inductance at resonance is:

$$V_L = X_L I = \omega L I = \frac{V}{R}$$

$$\therefore V_L = -\left(\frac{\omega L}{R}\right) \times V = QV \text{ volts, where}$$

$$Q_0 = \frac{\omega L}{R}$$

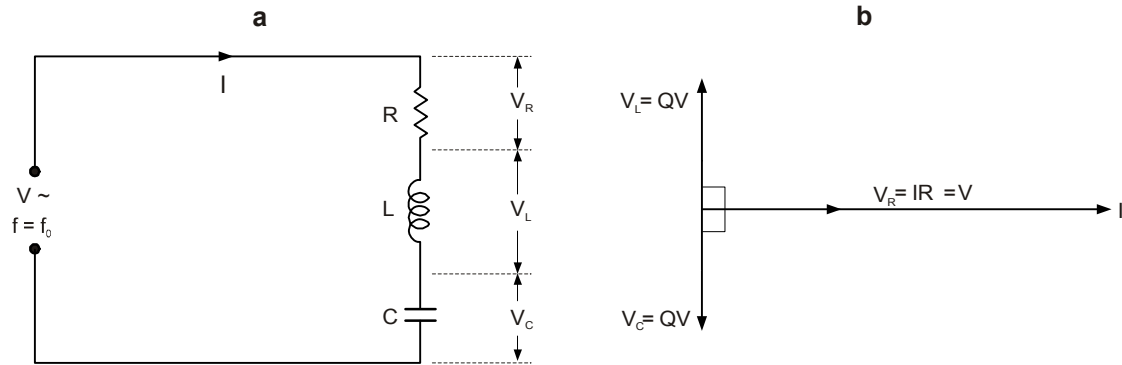
The voltage across the inductance is  $Q_0$  times the applied voltage  $V$  and voltage magnification has taken place.

c. **Capacitor.**  $Q_0$ , as already defined, is equal to  $\omega L/R$ . However, at resonance,  $\omega L = 1/\omega C$ . Hence  $Q_0$  is also equal to  $1/\omega CR$  and the voltage across the capacitance at resonance is:

$$V_C = X_C I = \frac{I}{\omega C} = \frac{1}{\omega C} \times \frac{V}{R}$$

$$\therefore V_C = \left(\frac{1}{\omega CR}\right) \times V = QV \text{ volts}$$

Thus the voltage across the capacitance at resonance is equal in magnitude but opposite in polarity to the voltage across the inductance (Fig 20b).

**14-8 Fig 20 Voltage Magnification**

38. When a series circuit is resonant to a given input, the voltage across either L or C at the input frequency can be many times greater than the input voltage. If  $V = 0.1V$  and  $Q = 100$ , the voltage across either L or C is 10V and this voltage can be applied to another circuit. A resonant series circuit is, therefore, a voltage magnifier.

## PARALLEL CIRCUITS

### Introduction

39. In this part, circuits consisting of different types of components connected in parallel are considered. The phase and magnitude relationships for currents and voltages are discussed, as they were for the series circuits. There are important differences in the construction of vector diagrams for series and for parallel AC circuits, and these are summarized as follows:

- a. Series Circuit.
  - (1) All components carry the same current in the same phase.
  - (2) The resultant applied voltage is the vector sum of the individual voltages across the separate components.
  - (3) Vector diagrams are constructed by drawing the voltage vectors relative to the current, which is the reference vector.
- b. Parallel Circuit.
  - (1) All branches have the same voltage across them in the same phase.
  - (2) The resultant supply current is the vector resultant of the individual currents in the separate branches.
  - (3) Vector diagrams are constructed by drawing the current vectors relative to the voltage, which is the reference vector.

### Resistance and Inductance in Parallel

40. The vector diagram for the circuit of Fig 21a is given in Fig 21b. It shows that:

- a. The applied voltage  $V$  is common to both R and L and is the reference vector.
- b. The current through the resistor is  $I_R = V/R$  amperes and is in phase with  $V$ . The vector  $I_R$  is drawn to any convenient scale in line with  $V$ .
- c. The current through the inductor is  $I_L = V/X_L$  amperes and is drawn to the same scale as  $I_R$ , lagging  $V$  by  $90^\circ$ .

d. The resultant supply current  $I$  is the vector sum of  $I_R$  and  $I_L$  and is found by completing the parallelogram and applying Pythagoras' theorem to Fig 21b. Thus:

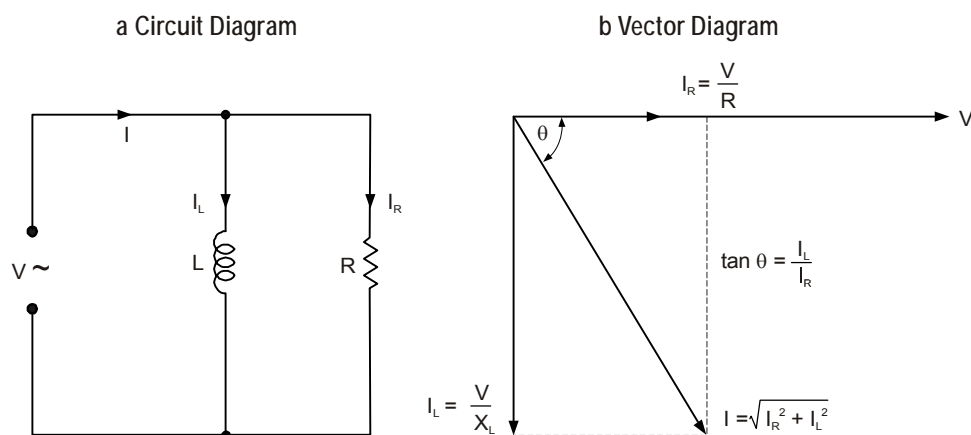
$$I = \sqrt{I_R^2 + I_L^2} = V \sqrt{\frac{1}{R^2} + \frac{1}{X_L^2}} \text{ amps}$$

e.  $I$  lags  $V$  by an angle  $\theta$ , where:

$$\tan \theta = \frac{I_L}{I_R} = \frac{V/X_L}{V/R} = \frac{R}{X_L} = \frac{\text{resistance}}{\text{reactance}}$$

f. From the equation in d,  $\frac{V}{I} = \frac{1}{\sqrt{\frac{1}{R^2} + \frac{1}{X_L^2}}} = \text{impedance (Z), ohms.}$

**14-8 Fig 21 R and L in Parallel**



### Resistance and Capacitance in Parallel

41. The vector diagram for the circuit of Fig 22a is given in Fig 22b. This vector diagram is constructed as follows:

a. The applied voltage  $V$  is the reference vector.

b. The current through the resistor is  $I_R = \frac{V}{R}$  amperes and is in phase with  $V$ .

c. The current through the capacitor is  $I_C = \frac{V}{X_C}$  amperes and leads  $90^\circ$  on  $V$ .

d. The resultant supply current  $I$  is the vector sum of  $I_R$  and  $I_C$  and is found by completing the parallelogram and applying Pythagoras' theorem to Fig 22b. Thus:

$$I = \sqrt{I_R^2 + I_C^2} = V \sqrt{\frac{1}{R^2} + \frac{1}{X_C^2}} \text{ amps.}$$

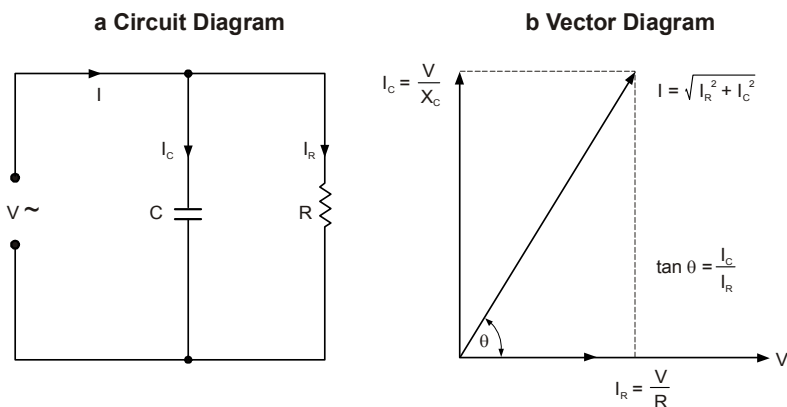
e.  $I$  leads  $V$  by an angle  $\theta$ , where:

$$\tan \theta = \frac{I_C}{I_R} = \frac{V/X_C}{V/R} = \frac{R}{X_C} = \frac{\text{resistance}}{\text{reactance}}.$$

f. From the equation in d,

$$\frac{V}{I} = \frac{1}{\sqrt{\frac{1}{R^2} + \frac{1}{X_C^2}}} = \text{impedance (Z), ohms.}$$

14-8 Fig 22 R and C in Parallel

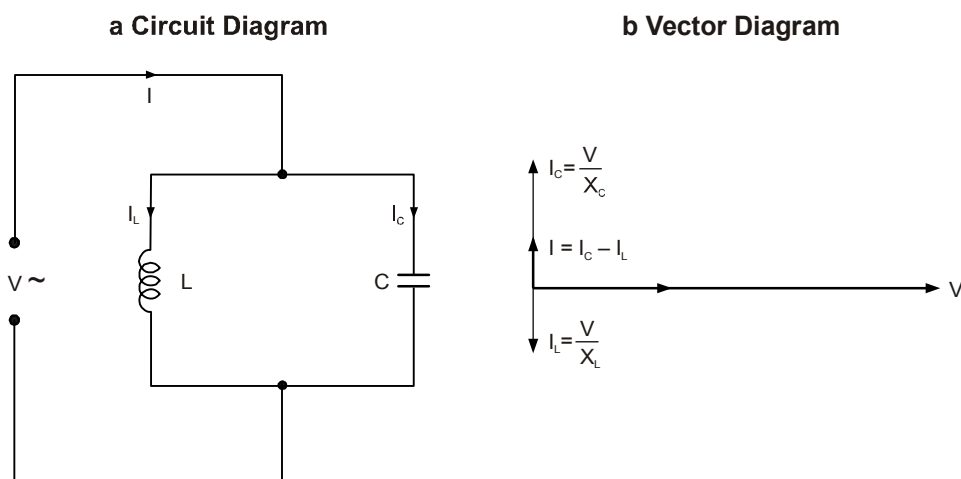


### Inductance and Capacitance in Parallel

42. Fig 23b is the vector diagram for the circuit of Fig 23a, and is constructed as follows:

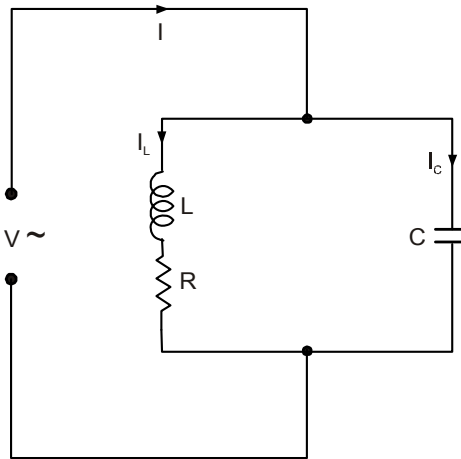
- The applied voltage  $V$  is the reference vector.
- The current through the inductance is  $I_L = V/X_L$  and lags  $V$  by  $90^\circ$ .
- The current through the capacitance is  $I_C = V/X_C$ , and leads  $V$  by  $90^\circ$ .
- The resultant supply current  $I$  is the vector sum of  $I_L$  and  $I_C$  and as  $I_L$  and  $I_C$  are  $180^\circ$  out of phase,  $I$  is therefore  $I_C \sim I_L$ .
- $I$  is  $90^\circ$  out of phase with  $V$ , leading or lagging  $V$  depending on the relative sizes of  $I_L$  and  $I_C$ . In Fig 23b  $I$  is shown leading  $V$  by  $90^\circ$  because  $I_C > I_L$ .

14-8 Fig 23 L and C in Parallel



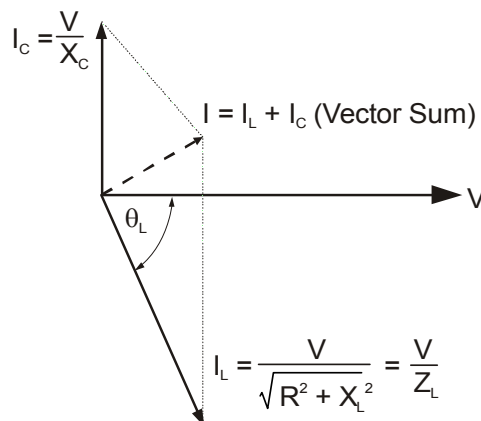
### Tuned Circuit with Resistance

43. Consider the circuit of Fig 24. This shows a practical arrangement of a coil and a capacitor in a parallel tuned circuit. The coil has certain power losses and these are represented by a resistance in series with the inductance of the coil. The capacitor losses are small in comparison and are ignored. This approximation is satisfactory for most purposes and is normally assumed when considering low power RF circuits.

**14-8 Fig 24 Practical Parallel Tuned Circuit**

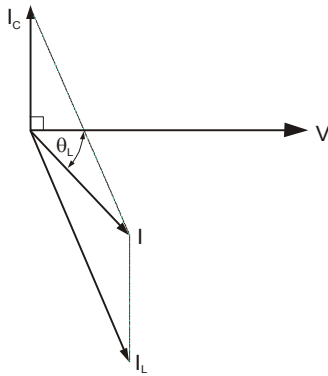
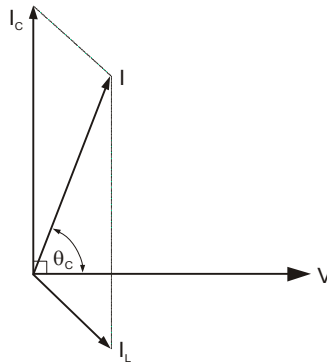
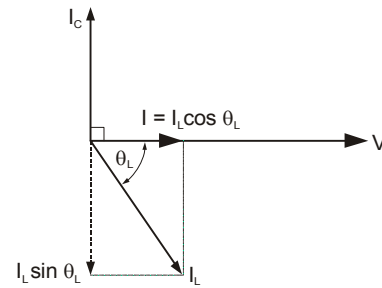
44. The vector diagram for the circuit is constructed as shown in Fig 25, where:

- The applied voltage  $V$  is the reference vector.
- The capacitive current is  $I_C = V/X_C$  amperes and leads  $V$  by  $90^\circ$ .  $I_C$  increases with frequency.
- The inductive current is  $I_L = V/\sqrt{R^2 + X_L^2}$ , and lags  $V$  by an angle  $\theta_L$  which is less than  $90^\circ$  by virtue of the series resistance.  $I_L$  decreases as the frequency increases.
- The resultant supply current  $I$  is the vector sum of  $I_C$  and  $I_L$ .

**14-8 Fig 25 Vector Relationships in a Parallel Tuned Circuit**

45. The phase of the supply current  $I$  relative to the applied voltage  $V$  depends on the supply frequency, as well as on the values of  $L$  and  $C$ . At a low frequency,  $I_C = V/X_C = V\omega C$  will be small and  $I_L = \frac{V}{\sqrt{R^2 + \omega^2 L^2}}$  will be large so that  $I$  lags  $V$  by an angle  $\theta_L$  (Fig 26a). At a high frequency,  $I_C$  will be large and  $I_L$  small, so that  $I$  leads  $V$  by an angle  $\theta_C$  (Fig 26b). At a particular frequency (the resonant frequency)  $I$  will be in phase with  $V$  (Fig 26c).



**14-8 Fig 26 Effect of Variation of Frequency****a Low Frequency****b High Frequency****c Resonant Frequency**

46. Resonance occurs in a parallel tuned circuit when the supply current  $I$  is in phase with the applied voltage  $V$ . For this to be so, the out-of-phase (or reactive) component of  $I_L$  must equal  $I_C$ . There is then no reactance offered by the circuit and its impedance is a pure resistance. Thus the condition for resonance in a parallel tuned circuit is that of zero reactance.

47. **Resonant Frequency.** The out-of-phase component of a current  $I_L$ , which is at an angle  $I_C$  to the reference, is  $I_L \sin \theta_L$  as shown in Fig 26c, and for resonance to occur  $I_L \sin \theta_L = I_C$ . It can be shown that at resonance, the resonant frequency is given by:  $f_o = \frac{1}{2\pi} \sqrt{\frac{1}{LC} - \frac{R^2}{L^2}}$ . Unlike in the series tuned circuit case,  $R$  has an effect on the resonant frequency of a parallel tuned circuit. In most cases, the resistance is very small and  $R^2/L^2$  can be ignored, so that for practical purposes the resonant frequency can be taken as  $f_o = \frac{1}{2\pi\sqrt{LC}}$  (Hz).

48. **Dynamic Impedance.** At resonance in a parallel tuned circuit there is zero reactance and the supply current  $I$  is in phase with the applied voltage  $V$ . From Fig 26c it is seen that at resonance,  $I$  is equal to the in-phase (or resistive) component of  $I_L$ . Thus at resonance  $I = I_L \cos \theta_L$  and it can be shown that  $Z = L/CR$ . Note that the smaller  $R$  is, the greater is  $Z$ ; if  $R$  is zero  $Z$  is infinite. The expression  $Z_D = L/CR$  is known as the dynamic impedance and is the purely resistive impedance of a parallel tuned circuit at resonance.

49. **Frequency at or Near Resonance.** The frequency at which resonance occurs is given by:

$$f_o = \frac{1}{2\pi} \sqrt{\frac{1}{LC} - \frac{R^2}{L^2}}$$

If the applied frequency is greater than  $f_o$  the current ( $I_C = V\omega C$ ) through the capacitor will increase whilst that through the inductor ( $I_L = V / \sqrt{R^2 + \omega^2 L^2}$ ) will decrease. If the frequency is lower than  $f_o$ , the reverse occurs.

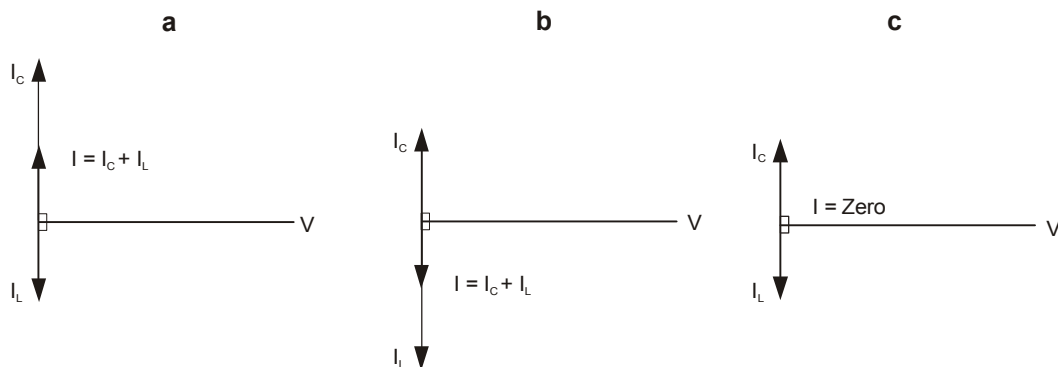
50. **Supply Current at or Near Resonance.** At resonance the supply current is in phase with the supply voltage and is at a minimum equalling  $VCR/L$ . When the frequency is above  $f_o$ , the supply current leads the supply voltage and the circuit is capacitive (see Fig 26b). When the frequency is below  $f_o$ , the circuit is inductive (see Fig 26a).

51. Impedance at or Near Resonance. At resonance the impedance equals  $L/CR$  and is at maximum. Either side of  $f_0$  the impedance falls by an amount dependent on:

- The departure from resonance.
- The ratio of  $C$  to  $L$ .
- The value of  $R$ .

52. **Resistance Ignored.** If the coil losses are so small that  $R$  can be ignored, the parallel tuned circuit has the 'ideal' form of Fig 23. The vector diagrams for this circuit will be as shown in Fig 27. If the applied frequency, or the values of  $L$  and  $C$ , are such that  $X_C$  is greater than  $X_L$ ,  $I$  leads  $V$  by  $90^\circ$  and the circuit behaves as a pure capacitance (Fig 27a). If  $X_L$  is greater than  $X_C$ ,  $I$  lags  $V$  by  $90^\circ$ , and the circuit behaves as a pure inductance (Fig 27b). If  $X_L$  equals  $X_C$ , the supply current into the circuit is zero and the circuit behaves as an infinite resistance (ie impedance is infinite). This is the resonant condition (Fig 27c), which occurs at a frequency  $f_0 = \frac{1}{2\pi\sqrt{LC}}$ .

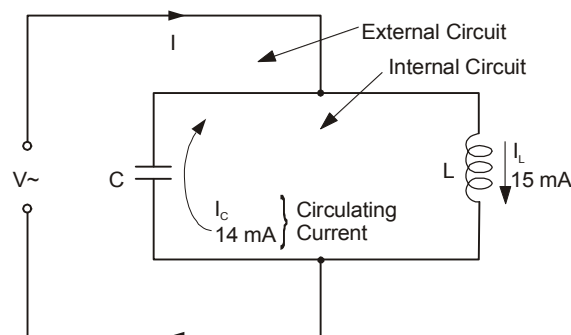
14-8 Fig 27 Effect of Varying  $X_C$  and  $X_L$



### Current Magnification

53. Although at resonance the current  $I$  into a parallel tuned circuit from the supply has a low value, this does not apply to the current within the closed LC loop. Thus the parallel tuned circuit may be thought of as having both an internal and an external circuit (Fig 28).

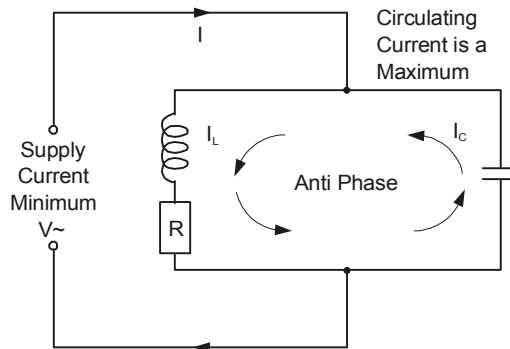
14-8 Fig 28 Supply and Circulating Current



54. In the ideal LC parallel circuit without resistance, the supply current  $I$  is the difference between  $I_L$  and  $I_C$ , since these currents are  $180^\circ$  out of phase. Thus  $I_L$  and  $I_C$  act in the same direction round the internal circuit, forming a circulating current. For example, if the ideal LC parallel circuit is near resonance,  $I_L$  may have a value of  $15 \text{ mA}$  and  $I_C$  a value of  $14 \text{ mA}$ . The circulating current is equal to the smaller of the two currents, and the supply current of  $1 \text{ mA}$  makes up the difference between  $I_L$  and  $I_C$ .

55. Fig 29 shows a practical parallel tuned circuit with resistance. At resonance,  $I$  has its minimum value of  $V/Z_D$  but the circulating current within the tuned circuit has a high maximum value equal to  $Q_0$  times the supply current. It can be shown that at resonance  $I_C = I_L = V \sqrt{\frac{C}{L}} = Q_0 I$ . Thus a parallel tuned circuit is a current magnifier, as distinct from the series tuned circuit which is a voltage magnifier.

**14-8 Fig 29 Current Magnification**



56. **Selectivity.** The selectivity of a parallel tuned circuit is discussed at greater length in Volume 14, Chapter 9. It may be noted here that:

- a. The higher the impedance at resonance in relation to the impedance at frequencies off resonance, the greater the selectivity of the circuit.
- b. A circuit with a high value of  $Q$  has high selectivity.

## CHAPTER 9 - SELECTIVITY OF TUNED CIRCUITS

Contents	Page
Introduction .....	1
SERIES TUNED CIRCUITS .....	1
General .....	1
The Effect of Resistance .....	2
The Effect of the L/C Ratio .....	2
Q Factor .....	3
$Q_0$ and Bandwidth .....	3
Effect of Supply Impedance .....	4
The Q Factor of Components .....	5
PARALLEL TUNED CIRCUITS .....	5
General .....	5
$Q_0$ and Selectivity .....	6
$Q_0$ and Bandwidth .....	6
Damping of Parallel Circuits .....	7
Effect of Supply Impedance .....	7

### Table of Figures

14-9 Fig 1 Variation of Current with Frequency - Series Tuned Circuit .....	2
14-9 Fig 2 Effect of Resistance on Selectivity .....	2
14-9 Fig 3 Effect of L/C Ratio on Selectivity .....	3
14-9 Fig 4 The Half-power Bandwidth .....	4
14-9 Fig 5 Effect of Supply Impedance on Selectivity .....	5
14-9 Fig 6 Variation of I and Z with Frequency about Resonance - Parallel Tuned Circuit .....	5
14-9 Fig 7 $Q_0$ and Selectivity .....	6
14-9 Fig 8 The Half - power Bandwidth .....	6
14-9 Fig 9 Effect of Damping on Selectivity .....	7
14-9 Fig 10 Effect of Supply Impedance on Selectivity .....	7

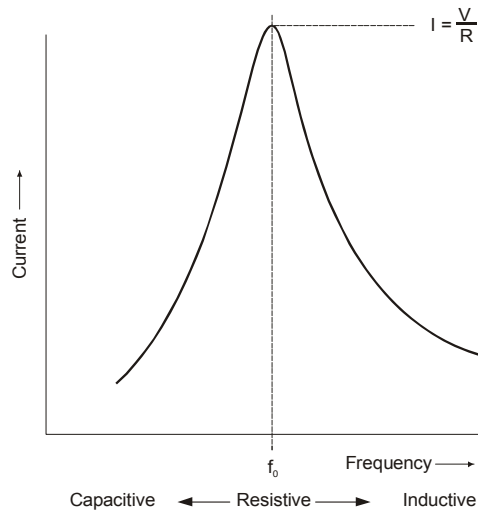
### Introduction

1. Selectivity of a tuned circuit is defined as the circuit's ability to pick out (select) a desired frequency, or band of frequencies, and reject the unwanted ones. The sharpness of response over a range of frequencies near resonance gives an indication of the selectivity of a circuit. In this Annex, considerations affecting the selectivity of series and parallel tuned circuits are examined.

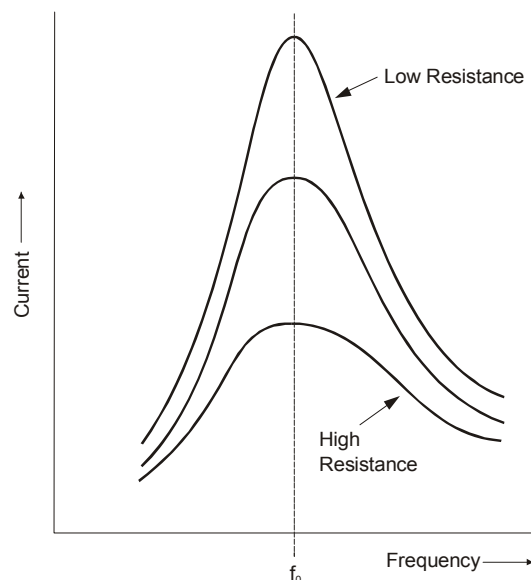
## SERIES TUNED CIRCUITS

### General

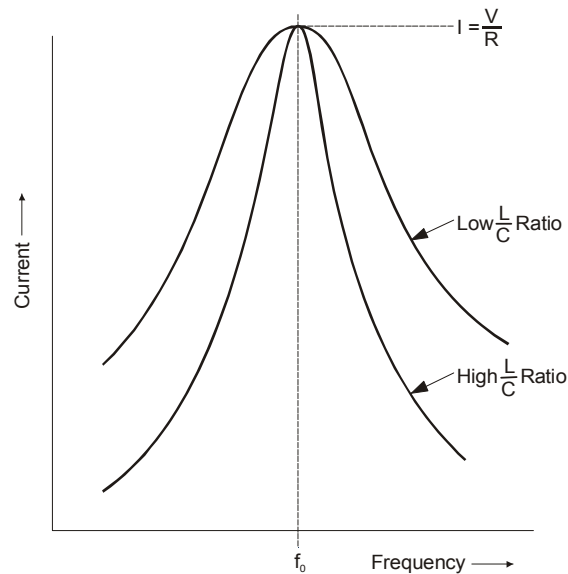
2. For the series tuned circuit, selectivity is a measure of the ease with which the circuit can accept an input at the resonant frequency as compared with inputs off resonance. Fig 19 of Volume 14, Chapter 8 shows that the impedance  $Z$  of a series tuned circuit falls as resonance is approached. At the resonant frequency, such a circuit allows the maximum current to flow through it, and is often termed an acceptor circuit. A graph showing the variation of current with frequency is at Fig 1.

**14-9 Fig 1 Variation of Current with Frequency - Series Tuned Circuit****The Effect of Resistance**

3. At resonance  $I = V/R$ , thus, if  $R$  is doubled, the current at resonance is halved. In a properly designed acceptor circuit, above resonance  $X_L$  is much greater than  $R$ , and below resonance,  $X_C$  is greater than  $R$ . Thus, the effect of  $R$  is progressively less important as the frequency is moved away from resonance. The effect of resistance is to reduce the current at frequencies near resonance to a far greater extent than at other frequencies, i.e. to 'flatten' the response curve as shown in Fig 2. Selectivity is therefore reduced by an increase in resistance.

**14-9 Fig 2 Effect of Resistance on Selectivity****The Effect of the L/C Ratio**

4. The resonant frequency of an acceptor circuit is given by  $f_0 = \frac{1}{2\pi\sqrt{LC}}$ . It is dependent on the product  $LC$ , thus, if the inductance of a circuit is doubled and the capacitance halved the resonant frequency remains unaltered. However, the ratio of  $L$  to  $C$  has been increased four times and this has an effect on the selectivity of the circuit. Fig 3 shows the response curves for two circuits having the same resonant frequency and equal values of resistance but different  $L/C$  ratios. From this it is seen that the greater the  $L/C$  ratio the more selective is the circuit.

**14-9 Fig 3 Effect of L/C Ratio on Selectivity****Q Factor**

5. The quantity used to represent the selectivity of a circuit at resonance is denoted by  $Q_0$  which is defined as the ratio of reactance to circuit resistance at the resonant frequency. It is usually considered for the coil, ie  $Q_0 = \frac{\omega_0 L}{R}$

$$\text{Thus, } Q_0 = \frac{2\pi f_0 L}{R}, \text{ but } f_0 = \frac{1}{2\pi\sqrt{LC}}$$

$$\therefore Q_0 = \frac{2\pi L}{R} \times \frac{1}{2\pi\sqrt{LC}} = \frac{L}{R} \times \frac{1}{\sqrt{LC}}$$

$$\therefore Q_0 = \frac{1}{R} \sqrt{\frac{L}{C}}$$

This expression supports paras 3 and 4 as it shows that the  $Q_0$  or selectivity of a tuned circuit is inversely proportional to  $R$ , and proportional to the ratio of  $L$  to  $C$ . Thus, a circuit with a high value of  $Q_0$  has high selectivity.

 **$Q_0$  and Bandwidth**

6. Series tuned circuits are used in radio to accept inputs at the resonant frequency and in the immediate neighbourhood of resonance. In order to present a high impedance to inputs considerably removed from resonance, a circuit must have a high  $Q_0$  value. In practice, values of  $Q_0$  vary from about 10 at audio frequencies to several hundreds at radio frequencies. The graph of current against frequency for a circuit having a high  $Q_0$  shows a sharp response curve (Fig 4). An alternative way of describing this is to say that the circuit has a narrow bandwidth. Bandwidth is defined as the separation between two frequencies either side of the resonant frequency at which the power has fallen to 50% of the maximum power. This is known as the half-power bandwidth. Since power is proportional to the square of the current, reducing the power to one-half means reducing the current by a factor of  $\sqrt{\frac{1}{2}} = 0.707$ . Thus in terms of current the bandwidth of a series tuned circuit is the difference between two frequencies  $f_1$  and  $f_2$  at which the current is 70% of the current at the resonant frequency  $f_0$  (Fig 4). The selectivity of a series

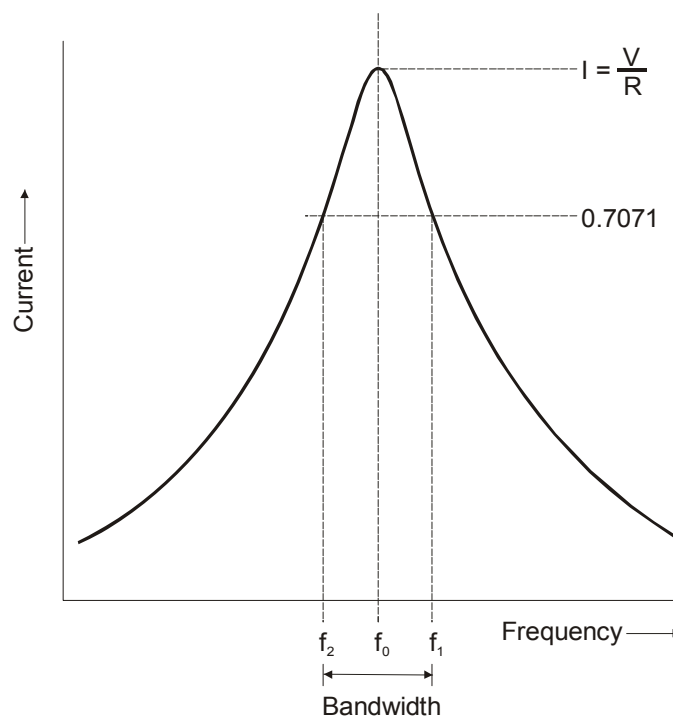
tuned circuit can therefore be defined either in terms of  $Q_0$  or in bandwidth. For purposes of calculation,  $Q_0$  and the bandwidth as defined above are related by the expressions:

$$Q_0 = \frac{\text{resonant frequency}}{\text{bandwidth}} = \frac{f_0}{f_1 - f_2}$$

Thus if the resonant frequency of a circuit is 200 kHz and the bandwidth required is 12 kHz,

$$Q_0 = \frac{200}{12} = 16.7$$

**14-9 Fig 4 The Half-power Bandwidth**

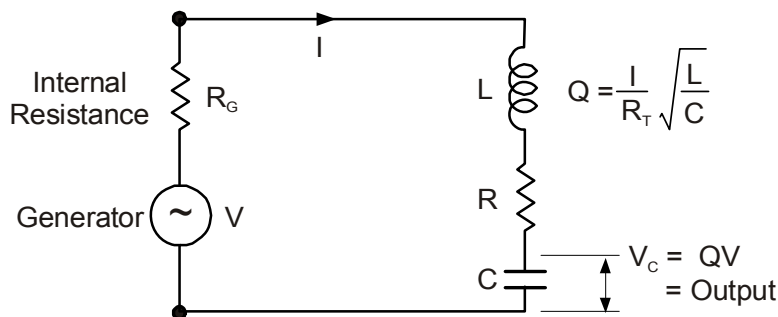


### Effect of Supply Impedance

7. When a series resonant circuit is used for tuning purposes, it is often intended that the voltage across the capacitor be applied to another stage. The voltage applied to the circuit itself can be considered as the supply from a generator with an internal resistance  $R_G$  (Fig 5). If  $R_G$  is considered non-reactive, it will modify the tuned circuit characteristics as follows: total circuit resistance is  $R_T = R + R_G$ , ie:

$$\text{effective } Q = \frac{1}{R + R_G} \sqrt{\frac{L}{C}}$$

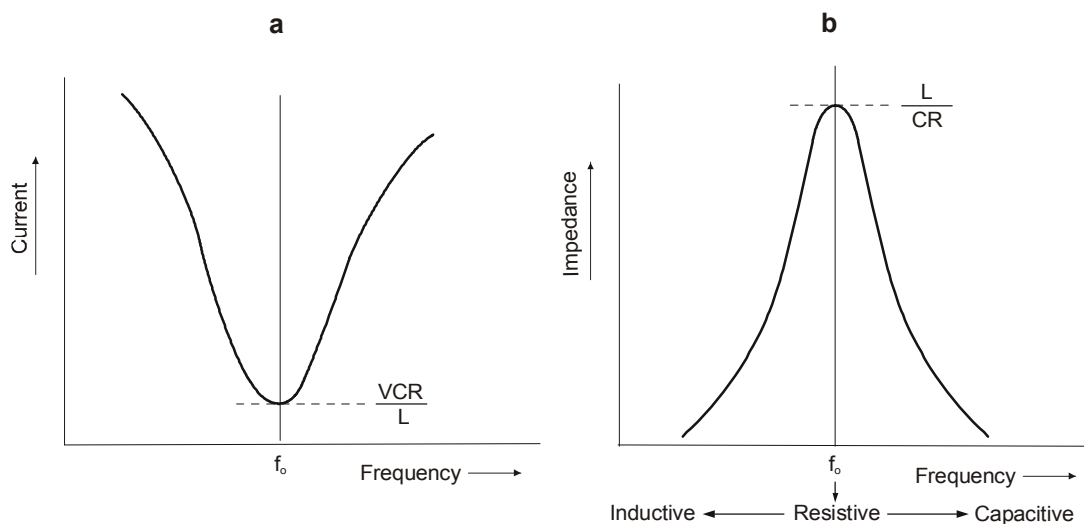
If  $R_G$  is large compared with  $R$ , then  $Q$  is reduced and the voltage across the capacitor at the resonant frequency ( $V_C = QV$ ) and the selectivity of the circuit are both relatively small. Thus, the series circuit is more selective when fed by a generator of low internal impedance.

**14-9 Fig 5 Effect of Supply Impedance on Selectivity****The Q Factor of Components**

8. It was shown in Volume 14, Chapter 8 that the power losses associated with inductors and capacitors can be expressed in terms of the power factor,  $\cos \theta = R/Z = \frac{R}{\sqrt{R^2 + X^2}}$  where  $R$  is the equivalent loss resistance of the component. In well-designed components,  $R$  is small compared with  $X$ , and the power factor approximates to  $R/X$  (i.e.,  $R/\omega L$  for inductors and  $\omega CR$  for capacitors). It is, however, more convenient to express the quality of inductors and capacitors in terms of the reciprocal of the power factor, namely  $\omega L/R = 1/\omega CR = Q$ . Thus, a coil having a high value of  $Q$  indicates a component with low losses.  $Q$  remains relatively constant with frequency since the value of loss resistance ( $R$ ) varies with frequency in much the same way as the reactance ( $X$ ). At audio frequencies,  $Q$  values for coils rarely exceed 10, whilst coils normally used for radio frequencies have  $Q$  values around 50 to 300.  $Q$  values from 100 to 300 are common with paper dielectric capacitors and from 1,000 to 3,000 for mica capacitors.

**PARALLEL TUNED CIRCUITS****General**

9. The variation of supply current with frequency for a parallel tuned circuit is shown in Fig 6a. The sharpness of the response curve denotes the selectivity of the circuit. For parallel circuits, selectivity is often defined in terms of impedance rather than of current, and the higher the impedance at resonance in relation to the impedance at frequencies off resonance, the greater the selectivity of the circuit. Fig 6b shows the variation of impedance with frequency.

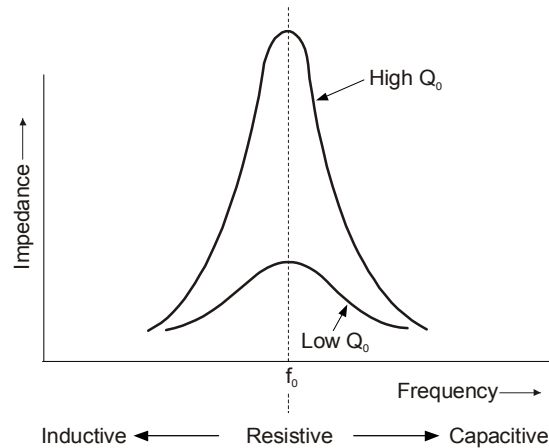
**14-9 Fig 6 Variation of  $I$  and  $Z$  with Frequency about Resonance - Parallel Tuned Circuit**



### $Q_0$ and Selectivity

10. An increase in resistance  $R$  will reduce the impedance at resonance ( $Z = L/CR$ ) to a greater extent than at frequencies removed from resonance. Since  $Q_0 = \frac{1}{R} \sqrt{\frac{L}{C}}$  is reduced, the response curve is 'flattened' and the circuit is made less selective. A variation in the ratio of  $L$  to  $C$  will also affect the impedance at resonance and, hence, the selectivity of the circuit. Thus, a circuit with a high value of  $Q_0$  has a high selectivity, and vice versa (Fig 7).

14-9 Fig 7  $Q_0$  and Selectivity

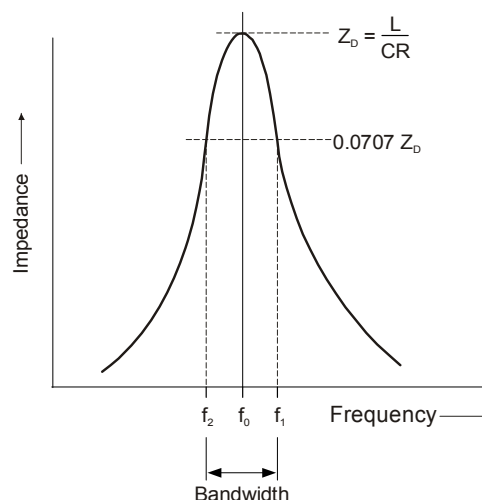


### $Q_0$ and Bandwidth

11. Parallel tuned circuits are used in radio equipment to 'reject' inputs at, and near, the resonant frequency, by offering a high impedance at those frequencies. To inputs at frequencies considerably removed from the resonant frequency, the parallel tuned circuit should offer a low impedance. To do this satisfactorily, the circuit must have a high  $Q_0$ , i.e. a narrow bandwidth. The bandwidth of a parallel tuned circuit is defined as the difference between two frequencies  $f_1$  and  $f_2$  at which the impedance has fallen to 70% of the resonant value (Fig 8). For purposes of calculations,  $Q_0$  and the bandwidth are related by the expression:

$$Q_0 = \frac{\text{resonant frequency}}{\text{bandwidth}} = \frac{f_0}{f_1 - f_2}$$

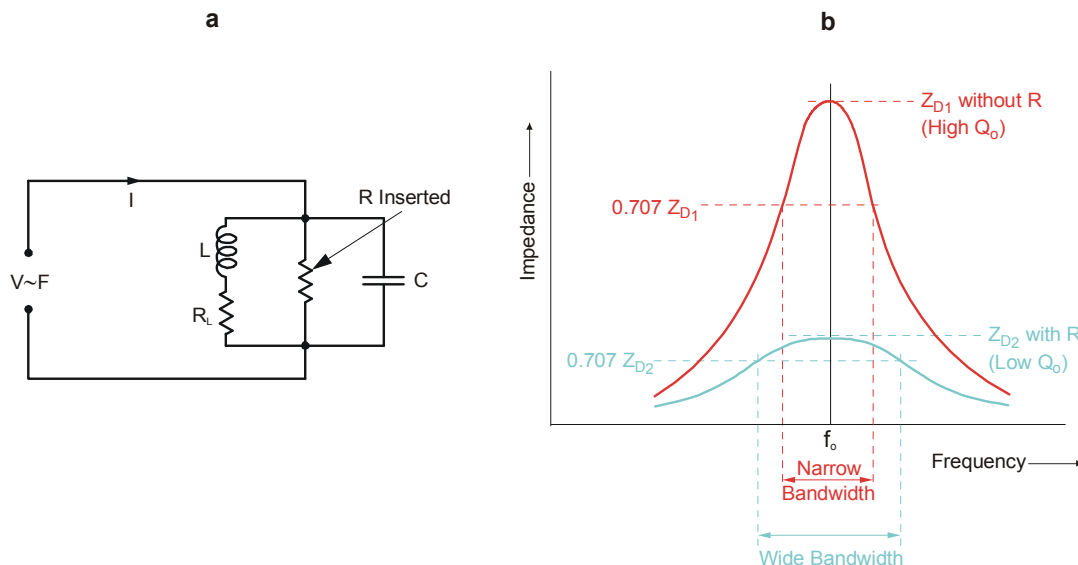
14-9 Fig 8 The Half - power Bandwidth



### Damping of Parallel Circuits

12. In some circuits in radio equipments, it is desired to 'pass' a wide band of frequencies in the neighbourhood of resonance, and for this a circuit with a wide bandwidth is required. From para 11, it is seen that the bandwidth of a parallel tuned circuit is inversely proportional to  $Q_0$  (ie bandwidth =  $f_0/Q_0$ ). Thus, a circuit with a low value of  $Q_0$  will give a wide bandwidth. To reduce  $Q_0$  an actual resistor is inserted in parallel with the rejecter circuit to 'damp' the response (Fig 9a). The impedance at resonance is then the dynamic impedance  $Z_D (= L/CR_L)$  in parallel with  $R$ . Consequently, the reduction of  $Q$  results in reduced impedance at resonance and a flattened response curve with a corresponding increase in bandwidth (Fig 9b).

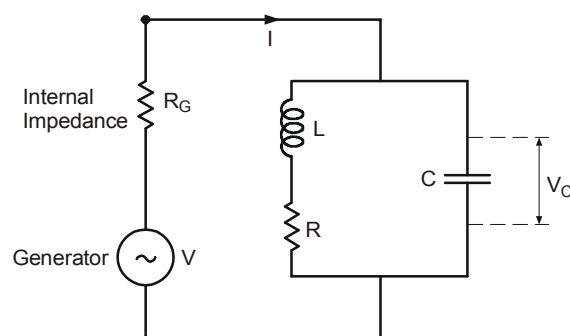
**14-9 Fig 9 Effect of Damping on Selectivity**



### Effect of Supply Impedance

13. When a parallel circuit is used for tuning purposes, the idea is that the voltage across the capacitor shall be applied to a further stage. The voltage applied to the circuit itself can be considered to be applied from a generator, and the behaviour of the circuit then depends on the impedance of the generator as well as all the characteristics of the circuit itself. If the internal impedance of the generator ( $R_G$ ) is small compared with the impedance of the parallel tuned circuit, then  $V_C$  will be approximately equal to  $V$  at all frequencies (Fig 10). If, however,  $R_G$  is large compared with the impedance of the circuit, then the current will be approximately equal to  $V/R_G$ . Thus,  $V_C$  will be proportional to the impedance of the parallel circuit and will vary with frequency, being a maximum at resonance. Hence, the parallel circuit is selective only if supplied from a generator having a high internal impedance.

**14-9 Fig 10 Effect of Supply Impedance on Selectivity**



## CHAPTER 10 - AC GENERATORS AND MOTORS

Contents	Page
POLYPHASE AC - GENERAL .....	2
Introduction .....	2
Effect of Load on Single-phase and Polyphase Supply .....	2
Generation of Polyphase Voltages .....	3
Advantages of Polyphase Systems .....	4
Symmetrical and Balanced Systems .....	4
AC GENERATORS .....	4
Single-phase Generators .....	4
Two-phase Generators .....	5
Three-phase Generators .....	5
INTERCONNECTION OF PHASES .....	6
General .....	6
Star (Y) Connection .....	6
Three-phase Four-wire Star Connection .....	7
Mesh or Delta ( $\Delta$ ) Connection .....	8
Power in Three-phase Circuits .....	8
AC MOTORS .....	9
Introduction .....	9
Production of a Rotating Magnetic Field .....	9
Synchronous Motors .....	10
Three-phase Induction Motors .....	12
Two-phase Induction Motors .....	15
Single-phase Induction Motors .....	15
Wound Induction Motors .....	16
Synchronous Induction Motors .....	17
Single-phase Commutator Motors .....	18
ROTARY INVERTERS AND CONVERTERS .....	18
Rotary Inverters .....	18
Rotary Converters .....	18

### Table of Figures

14-10 Fig 1 Power Variation with Inductive Load in a Single-phase Supply .....	2
14-10 Fig 2 Generation of Polyphase Voltages .....	3
14-10 Fig 3 Two-pole Single-phase AC Generator .....	5
14-10 Fig 4 Two-pole Three-phase AC Generator .....	5
14-10 Fig 5 Three-phase Star Connection .....	7
14-10 Fig 6 Three-phase Four-wire System .....	7
14-10 Fig 7 Three-phase Delta Connection .....	8
14-10 Fig 8 Three-phase Stator and Connections .....	9
14-10 Fig 9 Production of a Rotating Magnetic Field .....	10
14-10 Fig 10 Principle of Operation of a Synchronous Motor .....	11
14-10 Fig 11 Effect of Varying the DC Excitation to a Synchronous Motor .....	12
14-10 Fig 12 Three-phase Induction Motor .....	12
14-10 Fig 13 Principle of Induction Motor .....	13

14-10 Fig 14 Star-delta Starter .....	14
14-10 Fig 15 Torque-Speed Characteristics of an Induction Motor .....	14
14-10 Fig 16 Load-Speed Characteristics of an Induction Motor .....	15
14-10 Fig 17 Single-phase Induction Motor .....	16
14-10 Fig 18 Starting Device for a Single-phase Induction Motor .....	16
14-10 Fig 19 Slip Ring Induction Motor .....	17
14-10 Fig 20 Synchronous Induction Motor .....	17

## POLYPHASE AC - GENERAL

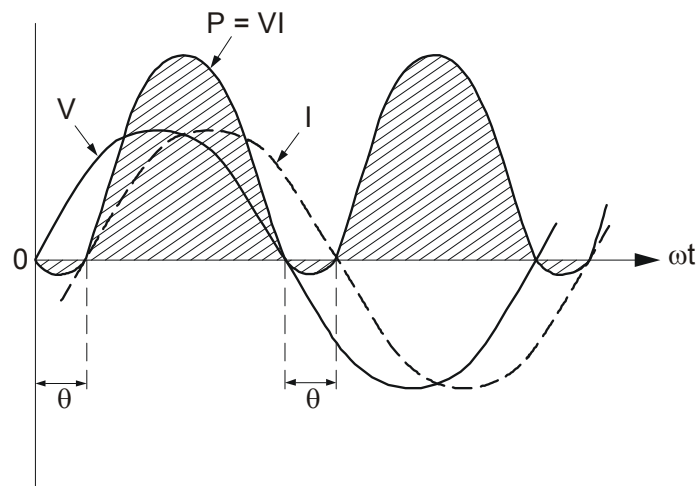
### Introduction

1. The simple AC generator (sometimes called an alternator), discussed in Volume 14, Chapter 3, produces a single-phase AC output at its slip rings. In other AC generators, the armature winding consists of two or more groups of series-connected coils, with their outer ends connected to separate slip rings. Two or more alternating voltages are then produced at the slip rings, these voltages being, in general, out of phase with each other. Such machines are known as polyphase or multiphase AC generators, or alternators.

### Effect of Load on Single-phase and Polyphase Supply

2. In polyphase systems, loads can be arranged to draw power from the generator at a uniform rate, so that the machine runs steadily under a uniform torque. Fig 1 shows the relationship between current and voltage when a single-phase AC generator is supplying an inductive load. The instantaneous power is the product  $VI$  at each instant, and is indicated by the shaded portion of the graph. When the power graph is above the time axis, it indicates that energy is being drawn from the generator; below this axis, energy is being returned from the inductive load. Since the load on the generator not only fluctuates during each half cycle, but also changes sign, the torque changes in a similar manner. Though this may happen in the individual cycles of each phase in a multiphase system, the power peak and zero values of various phases will not normally coincide because of the phase difference. Power is therefore drawn from the generator at a much more uniform rate, and the torque remains more steady.

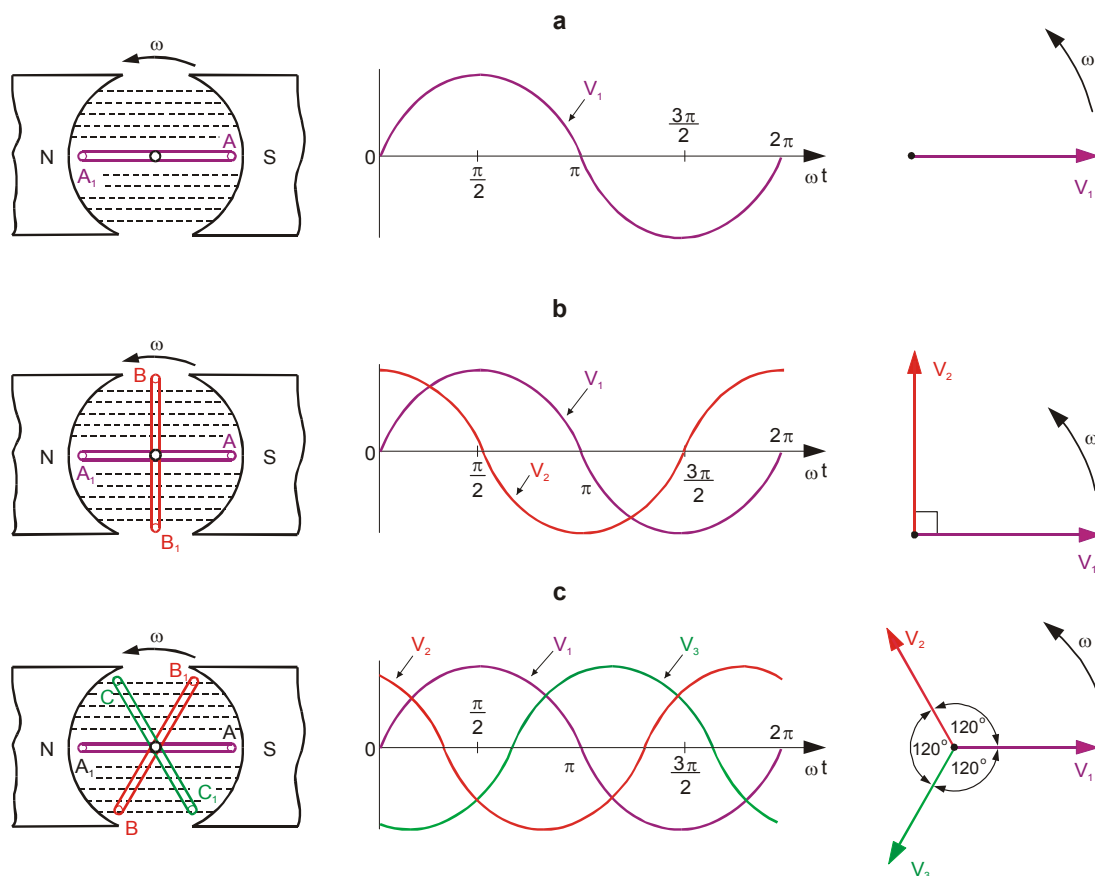
**14-10 Fig 1 Power Variation with Inductive Load in a Single-phase Supply**



### Generation of Polyphase Voltages

3. The output of a simple AC generator, consisting of a single group of coils rotating uniformly in a uniform magnetic field, is a sine wave ( $V_1$ ), as shown in Fig 2a, which represents the voltage in the coils ( $AA_1$ ) available at the slip rings.
4. If, on the same armature core, a second group of coils ( $BB_1$ ) is mounted at right angles to the first and connected to a second pair of slip rings, two independent voltages are available, differing in phase by  $90^\circ$ . This arrangement, shown in Fig 2b, represents a two-phase generator.
5. With three groups of coils ( $AA_1$ ,  $BB_1$  and  $CC_1$ ) wound independently on the same armature, and connected to three separate pairs of slip rings, three independent voltages are generated, as indicated in Fig 2c. The start of one coil is at an angle of  $120^\circ$  to the start of the next (i.e. A is  $120^\circ$  removed from B) and the three voltages differ in phase by  $120^\circ$ , the arrangement representing a three-phase generator. It should be noted that, when the voltage is in one phase at peak value, the voltage in each of the other phases is at half peak value in the opposite direction. The sum of the voltages in the three phases is thus zero; this holds good for all points in the cycle.

14-10 Fig 2 Generation of Polyphase Voltages



6. The armature can be wound with any number of symmetrically spaced groups of coils and independent pairs of slip rings; six-phase and twelve-phase AC supplies are sometimes used. However, since many of the advantages of polyphase voltages are available with three phases, the three-phase system is that in most general use for the generation and transmission of power.

## Advantages of Polyphase Systems

7. The advantages of polyphase systems can be summarized as follows:
  - a. Other parameters being unchanged, the power rating of a polyphase machine increases with the number of phases.
  - b. Both the heating loss for a given power transmitted, and the line voltage drop, are less than they would be if the whole power was transmitted by a single phase only.
  - c. Loads can be arranged to draw power from the generator at a uniform rate (see para 2).
  - d. Polyphase alternators can be made to work in parallel without much difficulty.
  - e. Polyphase motors may be self-starting, and provide uniform torque.

## Symmetrical and Balanced Systems

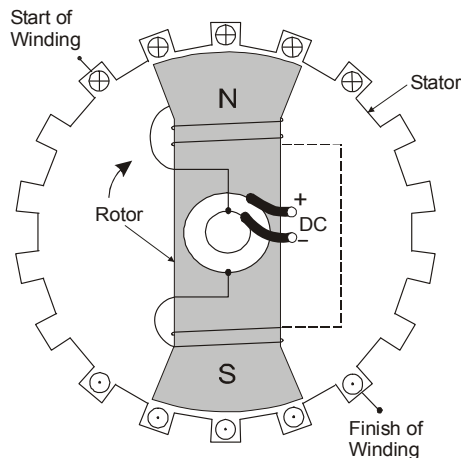
8. A polyphase system is said to be:
  - a. Symmetrical, when the phase voltages have the same amplitude and are displaced from one another by equal angles.
  - b. Balanced, when voltage, current and phase angle are the same for each phase. In a balanced three-phase system, the sum of the instantaneous values of voltage (or current) is zero.

# AC GENERATORS

## Single-phase Generators

9. A simple form of AC generator (the rotating armature type), consisting of a coil caused to rotate in a magnetic field by a prime mover, is described in Volume 14, Chapter 3. Most actual machines are of the rotating field type, designed the other way round. That is, the rotating part (rotor) consists of electromagnets energized from a DC supply, while the coils in which the generated emf is induced are wound on a fixed frame (stator). The two arrangements are electrically equivalent, since the relative motion of flux and coils that gives rise to an induced emf is the same in both cases. However, fixed connections to stationary windings are easier to insulate at high voltages than slip rings would be, and whether the output is single-phase or polyphase, only one pair of slip rings is needed for the relatively low voltage, DC energizing supply.

10. Fig 3 shows the arrangement of a two pole, single-phase AC generator of the rotating field type. The stator winding consists of a number of coils connected in series and inserted in slots cut in the inner surface of the laminated frame. The rotor is driven by a prime mover and carries the field windings, which are energized from a DC source via slip rings, as shown.

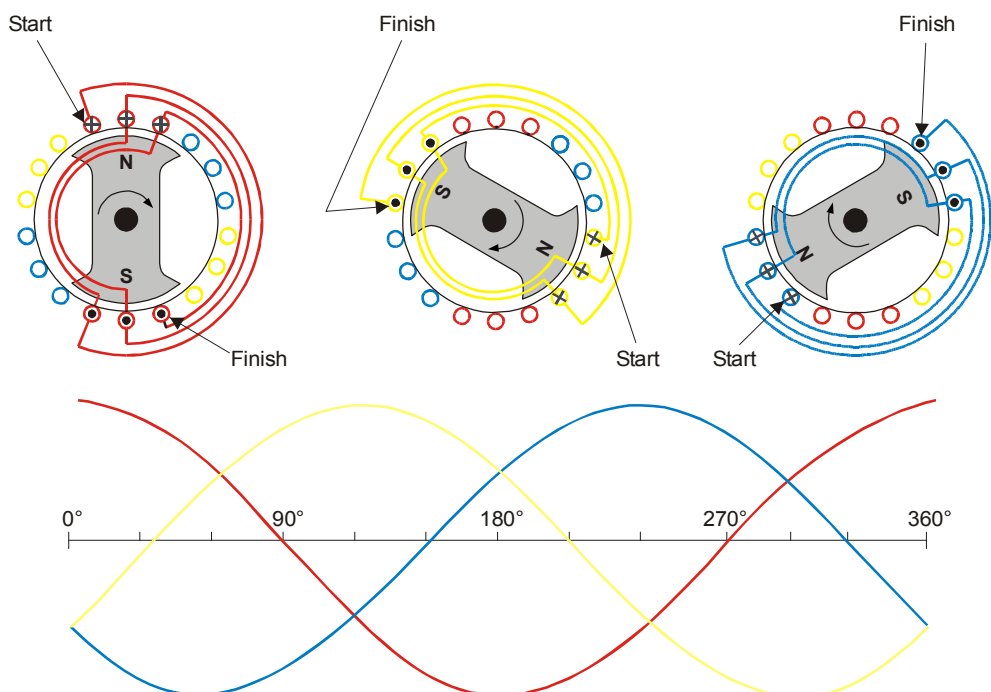
**14-10 Fig 3 Two-pole Single-phase AC Generator**

### Two-phase Generators

11. By mounting two separate coils on the rotor at right angles to each other, a two-phase output is produced. The voltages induced in each coil will be of the same magnitude and frequency, but there will be a  $90^\circ$  phase difference between the voltages. Two-phase supplies have limited application, mostly confined to direction control, where phase relationship determines the direction of rotation.

### Three-phase Generators

12. The stator of a three-phase AC generator has three separate windings equally spaced round the interior of the frame, and, between them, occupying all the slots in the laminations. The poles of the rotor sweep past each winding in turn, so that three alternating voltages are produced, differing in phase by  $120^\circ$ . This is shown in Fig 4. Standard practice identifies the phases by the colours red, yellow (or white), and blue; each phase then being referred to by its colour.

**14-10 Fig 4 Two-pole Three-phase AC Generator**

13. The emf generated in each conductor completes one cycle as it is passed by a pair of rotor poles. The machine shown in Fig 4 has two poles (one pair) on the rotor. However, some machines have many pairs of poles on the rotor, with the stator winding spaced accordingly, in a manner similar to that for the multi-pole DC generator discussed in Volume 14, Chapter 6. Thus, in one revolution of the rotor, the emf will complete a number of cycles corresponding to the number of pairs of poles in the generator. The frequency of the supply given by the machine will be:

$$\text{Frequency (F)} = \frac{p\eta_s}{60} \text{ Hz,}$$

where  $p$  = number of pairs of poles and  $\eta_s$  = speed in rpm.

The factor  $\eta_s$  is called the synchronous speed; it is the speed at which the machine must run in order to generate the required frequency. Thus:

$$\eta_s = \frac{f}{p} \times 60 \text{ (rpm)}$$

## INTERCONNECTION OF PHASES

### General

14. Each phase of a three-phase generator may be brought out to separate terminals and used to supply separate loads independently of one another. This method requires a pair of lines for each phase. The number of wires may be reduced, with a consequent saving in cable, if the phases are interconnected. There are two main methods of connecting the generator windings and the loads in three-phase systems:

- a. The Star, or Y, Connection. A development of this is the four wire star connection.
- b. The Delta, or Mesh, Connection.

### Star (Y) Connection

15. A star, or Y, connection is achieved by joining similar ends (the starting or the finishing ends) of the three windings to a common point (the neutral or star point). The other ends are joined to the line wires. In Fig 5, let  $E_a$ ,  $E_b$  and  $E_c$  be the phase voltages, and  $I_a$ ,  $I_b$  and  $I_c$  the phase currents, lagging behind their respective phase voltages by a constant angle  $\theta$  (ie a balanced load). The arrows indicate the assumed positive directions of the phase voltages. Voltage between lines is equal to vector difference between the two phase voltages. Hence, the voltage between lines 1 and 2 ( ${}_1V_2$ ) is obtained by the vector addition of vector  $E_a$  and vector  $E_b$  reversed. Thus:

$$\begin{aligned} {}_1V_2 &= E_a - E_b = 2E \cos 30^\circ \\ &= 2E \frac{\sqrt{3}}{2} \\ &= E\sqrt{3} \end{aligned}$$

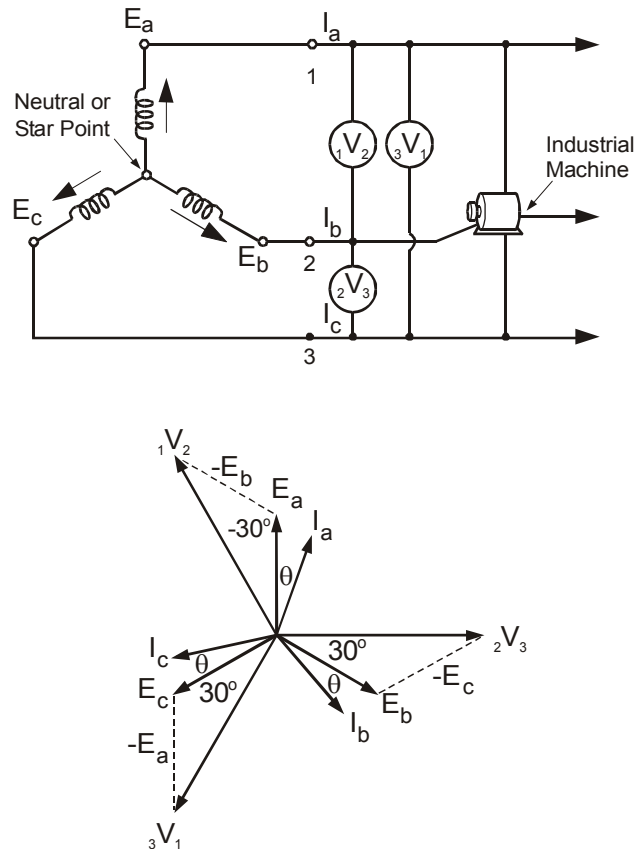
where the magnitudes of  $E_a = E_b = E_c = E$ , the phase voltage.

Similar results are obtained for  ${}_2V_3$ , and  ${}_3V_1$ , thus line voltages are each equal to  $\sqrt{3} \times$  phase voltage, or, generally:

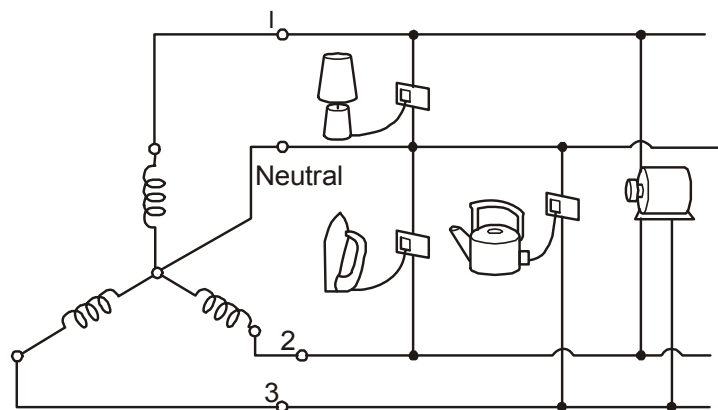
$$V_L = E\sqrt{3}$$

Since each line is in series with its individual phase winding, the line current must equal the phase current, i.e.  $I_L = I$ .



**14-10 Fig 5 Three-phase Star Connection****Three-phase Four-wire Star Connection**

16. In the three-phase four-wire star connection system, used in the National Grid, the phases of the alternator or transformer are connected in star, and three line wires taken from the terminals. In addition, a fourth wire is taken from the star point (Fig 6). This enables lighting loads and domestic services to be taken at phase voltage  $E$  (240 V) by connecting between line and neutral, and at the same time allows industrial machines to be supplied at line voltage  $\sqrt{3} E$  (415 V).

**14-10 Fig 6 Three-phase Four-wire System**

17. Three-phase motors constitute a balanced load, but domestic loads across the phases may not be balanced; in this case, the live conductors carry unequal currents, and the out of balance current is carried by the neutral.

### Mesh or Delta ( $\Delta$ ) Connection

18. In the Mesh, or Delta, system the phase windings are connected to form a closed mesh (Fig 7). Assuming, as before, a balanced load:

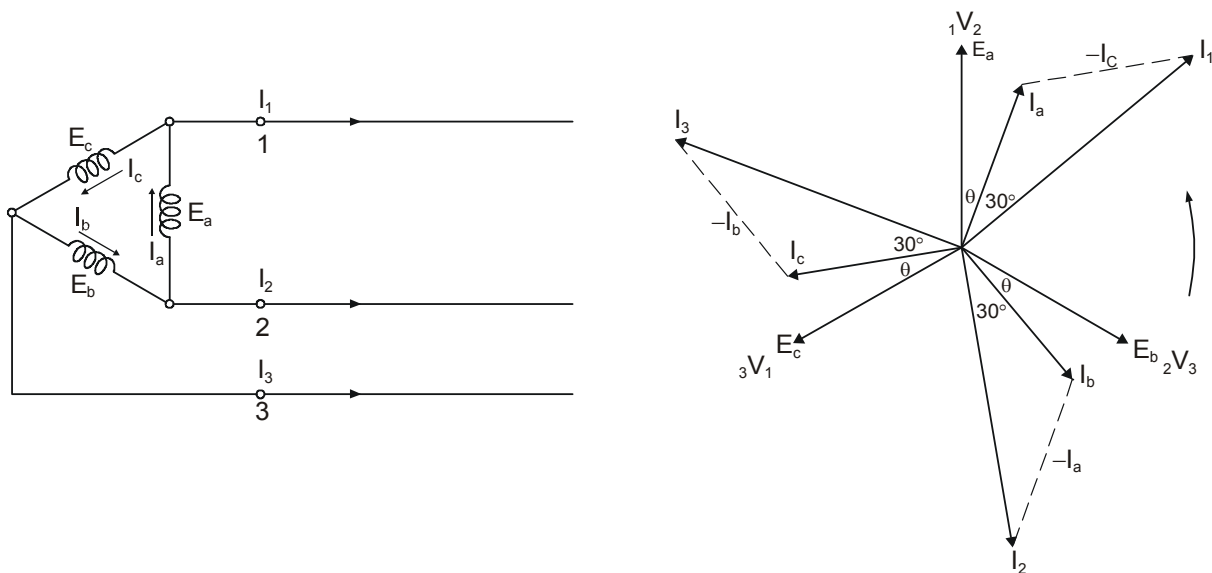
voltage between lines = phase voltage, i.e.  $V_L = E$ .

In this case, the line current is equal to the vector difference between two phase currents. Hence:

$$\begin{aligned} I_1 &= I_a - I_c = 2I \cos 30^\circ \\ &= I\sqrt{3} \end{aligned}$$

where the magnitudes of  $I_a = I_b = I_c = I$ , the phase current. Similar results are obtained for  $I_2$  and  $I_3$ , thus line currents are each equal to  $\sqrt{3} \times$  phase current, or generally,  $I_L = I\sqrt{3}$

14-10 Fig 7 Three-phase Delta Connection



### Power in Three-phase Circuits

19. In a balanced three-phase circuit, the total power equals three times the power per phase, ie

total power =  $3 \times$  phase voltage  $\times$  phase current  $\times$  power factor,

$$\text{or, } P = 3VI \cos \theta \text{ watts,}$$

where  $\theta$  is the angle between phase voltage and phase current.

Expressing this in line quantities:

a. **Star Connection.**

$$P = 3 \frac{V_L}{\sqrt{3}} I_L \cos \theta = \sqrt{3} V_L I_L \cos \theta$$

b. **Delta Connection.**

$$P = 3V_L \frac{I_L}{\sqrt{3}} \cos \theta = \sqrt{3} V_L I_L \cos \theta, \text{ as for star connection.}$$

## AC MOTORS

### Introduction

20. There are three main types of AC motor - the synchronous motor, the induction motor, and the series or commutator motor. Of these, the induction motor is the most common in industry; the commutator motor is the most common in the home.

21. As most large AC motors do not have commutators (some do not even have slip rings), they are more trouble-free in operation than DC motors. Some are also useful as constant speed motors, their speed being determined by the AC supply frequency.

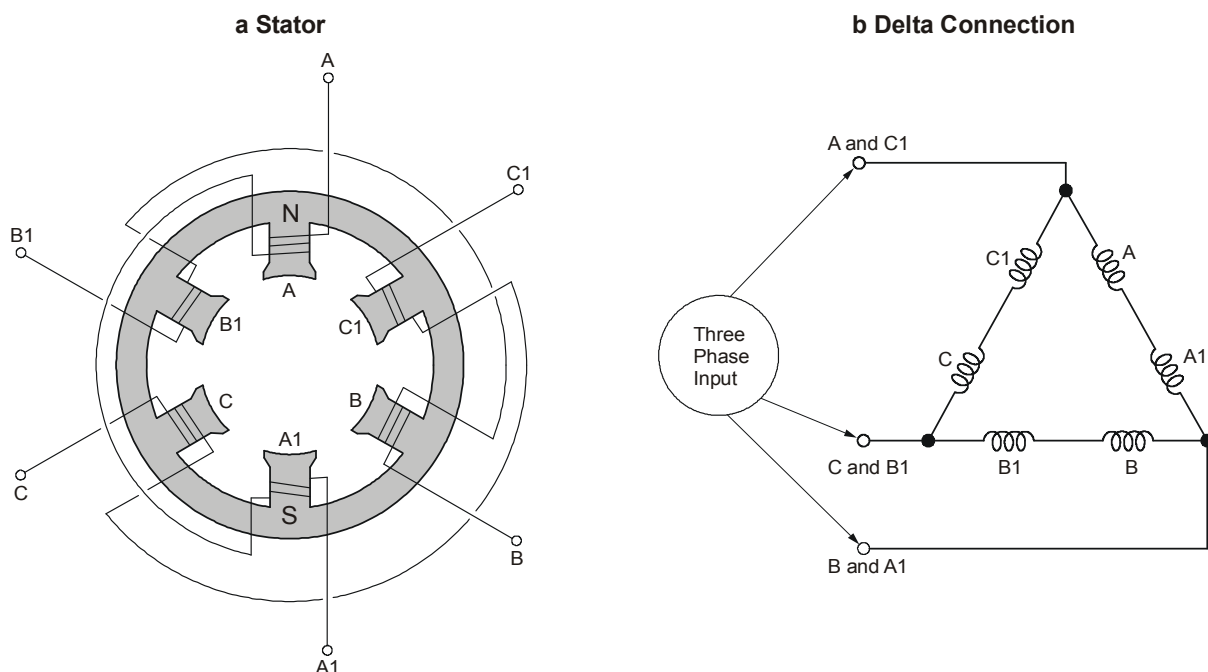
22. AC motors may be operated from single-phase or polyphase supplies; the basic principles of operation are similar in both cases. Synchronous and induction AC motors work on the principle that AC applied to the stator produces a rotating magnetic field, which causes the rotor of the machine to turn with the field.

### Production of a Rotating Magnetic Field

23. A rotating magnetic field can be produced by applying a three-phase supply to a stationary group of coils, provided the latter are suitably wound, spaced, and connected. The field produced is of unvarying strength, and its speed of rotation is directly related to the frequency of the supply.

24. Fig 8a shows a typical three-phase stator. The two windings in each phase (eg A and A<sub>1</sub>) are connected in series, and are so wound that current flowing through the two windings produces a North pole at one of them and a South pole at the other. Thus, if current is flowing in the A phase in the direction from the A to the A<sub>1</sub> terminal, pole-piece A becomes a North pole and A<sub>1</sub> a South pole. In operation, the three-phase stator is connected in delta, as shown in Fig 8b, so that only three terminals, each common to two of the windings, are provided for the three-phase input.

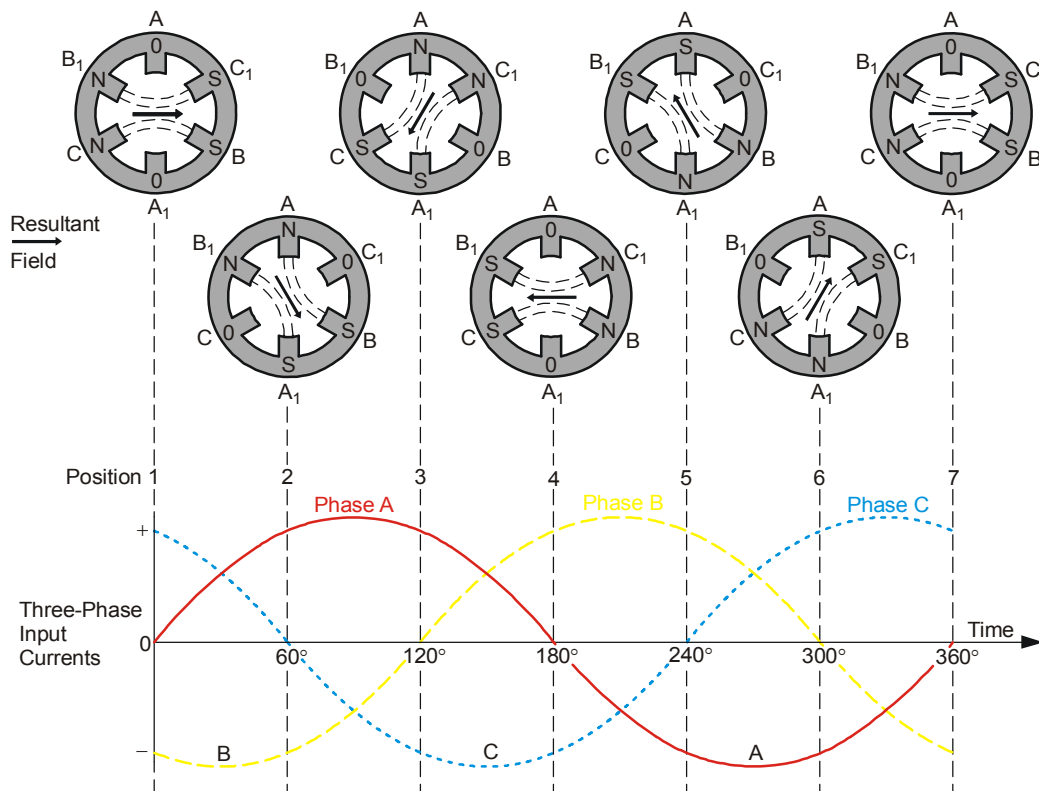
14-10 Fig 8 Three-phase Stator and Connections



25. At any instant, the magnetic field generated by one particular phase is proportional to the current in that phase; as the current alternates, so does the magnetic field. As the currents in the three phases are  $120^\circ$  out of phase, the three magnetic fields will also alternate  $120^\circ$  out of phase with each other, and the resultant magnetic field is the vector sum of these three.

26. Fig 9 shows how the magnetic fields add up to give a resultant magnetic field which continuously shifts in direction. After one complete cycle of AC input, the resultant magnetic field has shifted through  $360^\circ$ , or one revolution. Thus, although the coils are stationary, the application of three-phase AC produces a magnetic field that rotates at the frequency of the supply.

**14-10 Fig 9 Production of a Rotating Magnetic Field**



27. At position 1 in Fig 9, the current in phase A is zero, as shown by the graph; the current in phase C is positive and flows in the direction C to C<sub>1</sub>, and that in phase B is negative and flows in the direction B<sub>1</sub> to B. Equal currents therefore flow in opposite directions through the B and C phase windings, and magnetic poles are established as shown. Since the magnetic fields of the B and C phases are equal in amplitude (equal currents), the resultant field lies in the direction shown by the arrow.

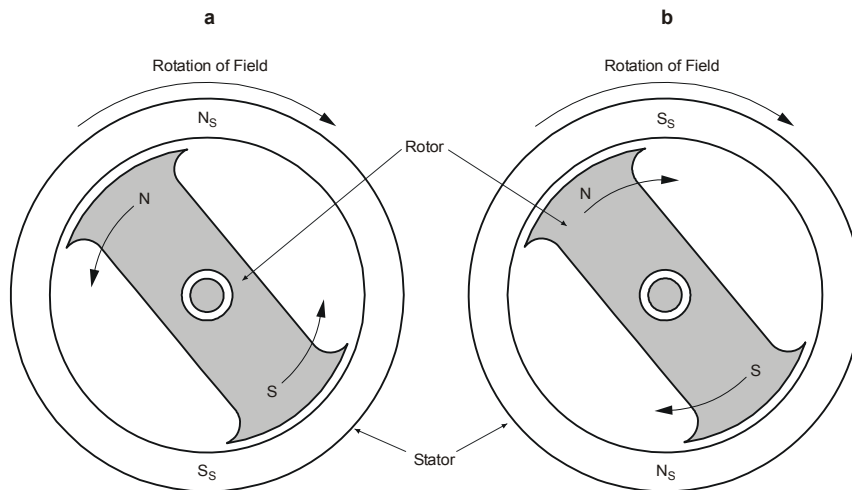
28. Position 2 shows the condition when the supply cycle has advanced  $60^\circ$ . The current C is now zero; A is positive and B negative. The resultant field is as shown. The other positions shown are at intervals of  $60^\circ$ . Thus, the field rotates one complete revolution during one complete cycle of the AC supply. An input frequency of 50 Hz produces a field rotating at 50 revolutions per second, i.e. at 3,000 rpm.

### Synchronous Motors

29. The AC generator, like the DC generator, is a reversible machine; if supplied with electrical energy, it will run as a motor. An AC generator used in this manner operates as a synchronous motor (so called, because the speed of the machine is dependent upon the synchronous speed of the rotating field produced by the input to the stator windings).

30. When a three-phase stator winding is supplied with three-phase AC, a magnetic field, of constant magnitude and rotating at synchronous speed, is produced within the stator gap. In a two-pole stator supplied at 50 Hz, this is equivalent to two poles ( $N_s$  and  $S_s$ ) rotating at 3,000 rpm. The rotor carries the field windings, which are supplied with DC to produce two poles (N and S). With the rotor stationary in the position shown in Fig 10a, there will be repulsion between N and  $N_s$ , and between S and  $S_s$ . A torque will therefore be exerted on the rotor in an anti-clockwise direction.

**14-10 Fig 10 Principle of Operation of a Synchronous Motor**



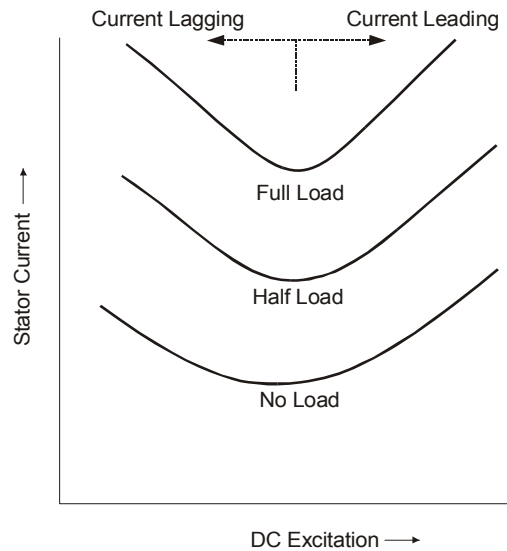
31. Half a cycle (i.e. 0.01 sec) later, the poles of the rotating stator field will have changed position, as shown in Fig 10b. There is now attraction between N and  $S_s$ , and between S and  $N_s$ , so that a clockwise torque is applied to the rotor. Because of its inertia, the rotor cannot respond to this rapidly alternating torque, and so remains at rest. The synchronous motor cannot, therefore, be started simply by applying AC to the stator winding, even if the rotor is already supplied with DC.

32. Suppose now that the rotor, driven by an external force, is already turning in a clockwise direction at just below synchronous speed. The relative speed between the rotating field and the rotor will be low, and eventually the N pole of the rotor will become adjacent to the  $S_s$  pole of the rotating field, as the latter overtakes the rotor. At this point, the two magnetic fields will lock together, and the rotor will maintain its position relative to the rotating field; that is, it will rotate at synchronous speed. The synchronous motor is usually started and run up towards the synchronous speed with the help of a small induction motor.

### 33. Characteristics.

a. **Effect of Load.** When a mechanical load is applied to a synchronous motor, the electrical input to the motor increases, the speed remaining constant. If too great a load is applied, the machine is pulled out of synchronism and comes to rest. The torque at which this occurs is called the 'pull-out torque'.

b. **Effect of DC Excitation.** If the mechanical load is kept constant, and the DC excitation varied, the back emf varies, and hence the supply current to the stator varies. A graph showing the variation of stator current as the excitation is varied is shown in Fig 11. The greater the load on the machine, the greater is the current and the higher is the curve on the graph. At a low value of DC excitation, the stator current is large and lags the applied voltage. At normal excitation, the current is a minimum and the phase angle zero. At a high value of excitation, the current again increases, but it now leads the applied voltage. This is an important property of the synchronous motor since, by over-exciting the field, the machine is made to take a leading current which compensates for any lagging current taken by other apparatus connected to the same supply. The power factor of the supply is thereby improved.

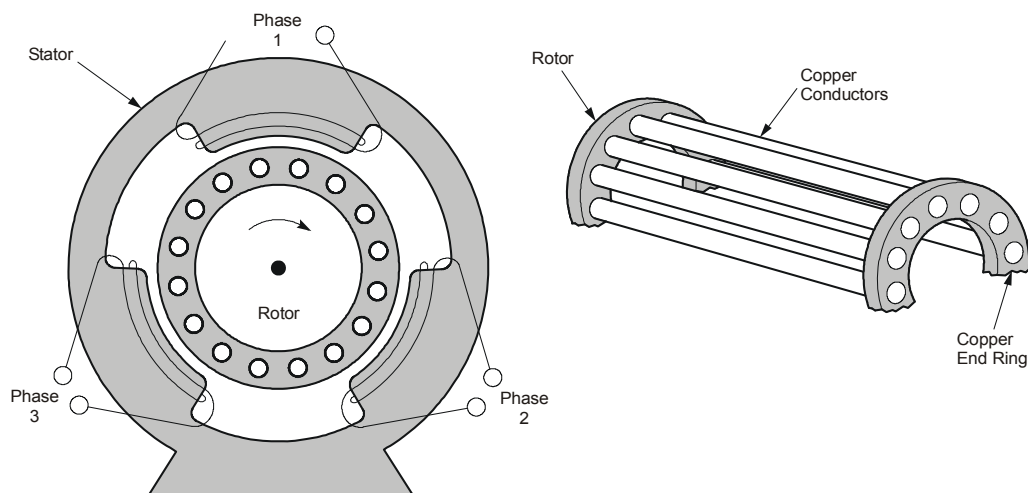
**14-10 Fig 11 Effect of Varying the DC Excitation to a Synchronous Motor**

34. **Summary.** The synchronous motor:

- a. Requires to be started by an external prime mover.
- b. Runs only at the synchronous speed.
- c. Can be used to adjust the power factor of a system at the same time as it is driving a mechanical load.

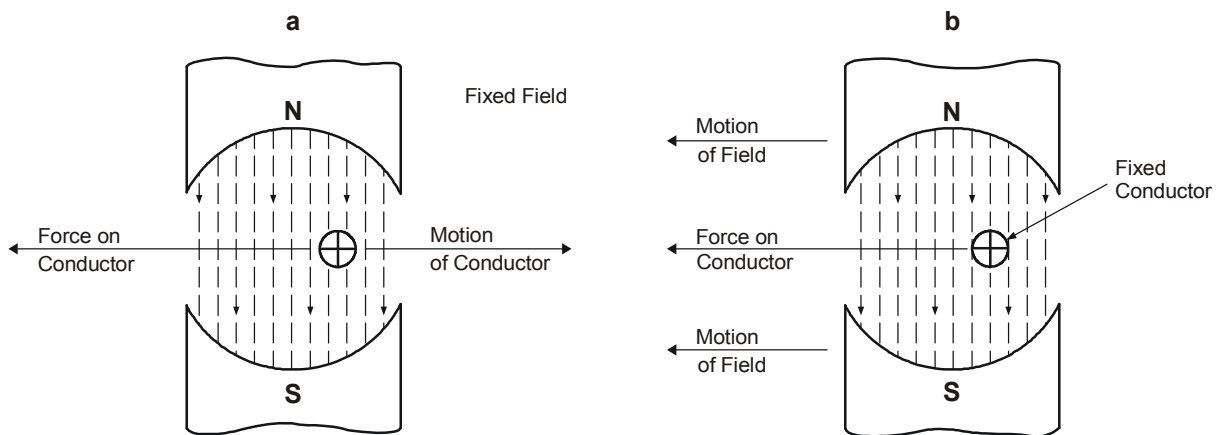
### Three-phase Induction Motors

35. **Construction.** Fig 12 shows the basic construction of a simple type of three-phase induction motor (note that the stator windings are usually in pairs of poles as in Fig 8a). It consists of a three-phase stator winding supplied with three-phase AC to produce a rotating magnetic field. The rotor consists of a set of stout copper conductors, laid in slots in a soft-iron armature, and welded to copper end rings, thus forming a closed circuit. This is termed a 'squirrel-cage' rotor and it will be seen that there is no electrical connection to the rotor.

**14-10 Fig 12 Three-phase Induction Motor**

36. **Principle.** If a conductor is set at right angles to a magnetic field, as in Fig 13a, and moved across the flux from left to right, the direction of the induced emf will be into the paper (Fleming's Right-hand Rule – Volume 14, Chapter 6). If this conductor is part of a complete circuit, a current will be established in the direction of this voltage, and there will be a force on the conductor tending to urge it from right to left (Fleming's Left-hand Rule – Volume 14, Chapter 6). The same relative motion of field and conductor applies if the conductor is stationary and the field moves from right to left, as in Fig 13b; if the circuit is completed so that current can be established in the conductor, the conductor tends to move from right to left. The conductor experiences a force moving it in the same direction as the field's motion. As the conductor follows the field, the relative motion is reduced, thereby reducing the conductor current and the force on the conductor. Thus, the conductor speed is limited to something less than that of the field, otherwise there will be no relative motion, no current, and no torque.

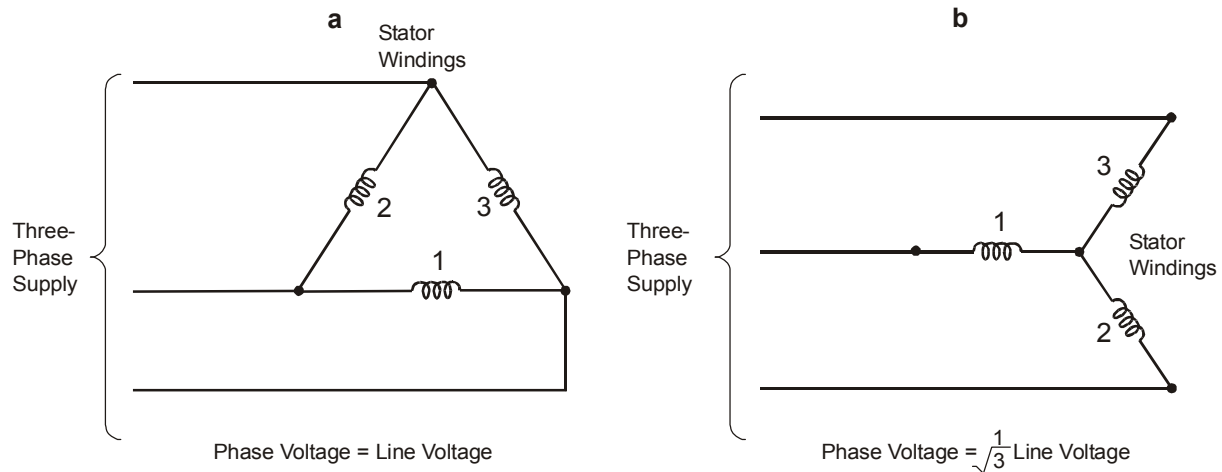
14-10 Fig 13 Principle of Induction Motor



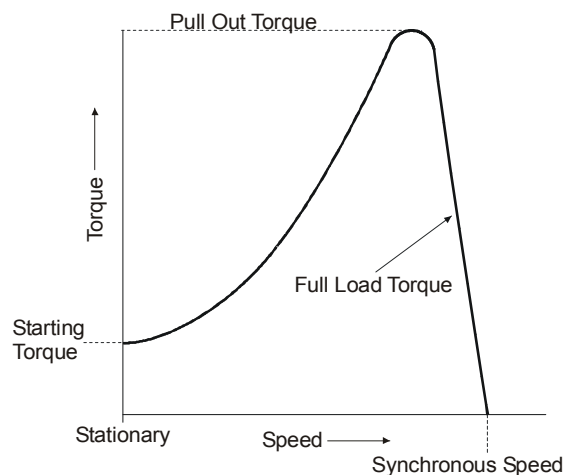
37. The squirrel-cage rotor of the induction motor, set in the rotating field of the stator, should accelerate until it is running steadily at a speed which is slightly less than the synchronous speed at which the magnetic field rotates. Thus, the rotor runs slightly slower than the rotating field, the amount depending on the load: the larger the load the greater the speed difference. In practice, very little speed change occurs between a light and a heavy load, and the main use of an induction motor is to drive a load at relatively constant speed.

38. **Slip.** The difference between the synchronous speed and the rotor speed, measured in rpm, is termed the 'slip speed'. The ratio of slip speed to synchronous speed, expressed as a percentage, is termed 'slip'. For example, a six-pole motor supplied at 50 Hz will have a synchronous speed of  $n_s = 60f/p = 60 \times 50/3 = 1,000$  rpm. With a rotor speed of 960 rpm, the slip speed will be  $1,000 - 960 = 40$  rpm and the slip will be  $40 \times 100/1,000 = 4\%$ . In practice, the value of slip varies from 2%, for large machines, to 6% for small machines.

39. **Starting Current.** When the stator winding is energized and the rotor stopped, the slip is 100% and maximum emf is induced in the rotor. A heavy current is thus established in the rotor, and this produces a flux which opposes and weakens the stator flux. The self-induced emf in the stator is therefore reduced, and a heavy current is taken by the stator winding on starting. To reduce this heavy starting current, the voltage applied to the stator windings should be at a reduced level on starting, until the rotor is turning at such a speed that its effect on the stator current is negligible. The normal way of doing this is to use a 'star-delta' starting switch. For normal running, the motor is designed to operate with the stator phases mesh or delta connected to the supply via the switch (Fig 14a), so that the phase voltage is equal to the line voltage. For starting, the stator windings are connected up in star (Fig 14b) via the switch to the supply, so that the phase voltage is  $\frac{1}{\sqrt{3}}$  of the normal voltage. This reduced voltage limits the starting current to a safe level.

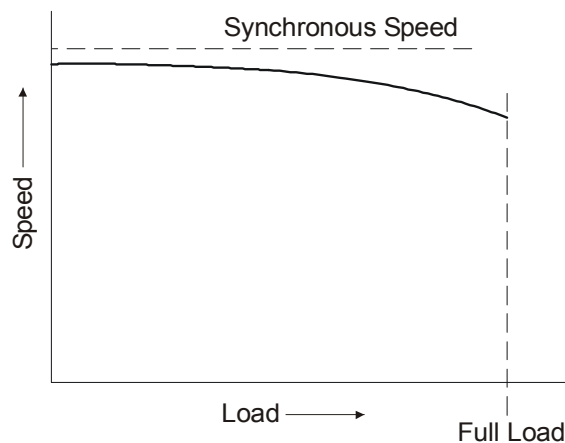
**14-10 Fig 14 Star-delta Starter**

40. **Torque.** The frequency of the current induced in the rotor is the frequency with which the stator field rotates relative to each conductor. When the rotor is at rest, this frequency equals the supply frequency. When the motor is running lightly loaded, the slip is small and the frequency of the induced rotor current may be only a few cycles per second. However, the resistance of a squirrel-cage rotor is small and its inductance high. Its impedance will, therefore, be large at the frequency of the supply when the rotor is stationary, and much less when it is running. Thus, on starting, the rotor current and the rotor emf are nearly  $90^\circ$  out of phase. The flux produced by this lagging rotor current is such that there is little interaction between it and the stator flux, and the starting torque is poor. As the rotor current comes round into phase with the rotor emf with increased rotor speed (decreased slip and inductive reactance), the rotor and stator fluxes come more into phase, and the torque increases. A typical torque-speed characteristic curve is shown in Fig 15.

**14-10 Fig 15 Torque-Speed Characteristics of an Induction Motor**

41. **Load-Speed Characteristics.** Off load, the torque developed by the rotor is only that required to overcome friction, and, in this condition, the rotor speed is almost synchronous. When a load is applied, the rotor slows down to the point at which the resultant increased driving torque balances the load torque. The fall in speed from no load to full load is small, as shown in Fig 16. As the load is increased, slip increases, until eventually a load is reached at which any further increase in slip, instead of increasing torque, reduces torque, and the motor will shut down. Usually, therefore, motors are designed so that the pull-out torque is at least twice the normal full load torque, to give an ample operating margin.



**14-10 Fig 16 Load-Speed Characteristics of an Induction Motor**

42. The speed of a squirrel-cage motor is not easily controlled, since it is related to synchronous speed, and its main use is on devices where a fairly constant speed is required. One typical use is as the prime mover for generators used in control systems, where ratings of 2 to 30 hp at speeds of up to 3,000 rpm may be required. The squirrel-cage motor can be readily adapted for frequent reversing. To do this, it is necessary to reverse the direction of rotation of the stator field, and this is achieved by changing over any two of the three connections at the stator terminals.

43. **Summary.** The three-phase squirrel-cage motor:

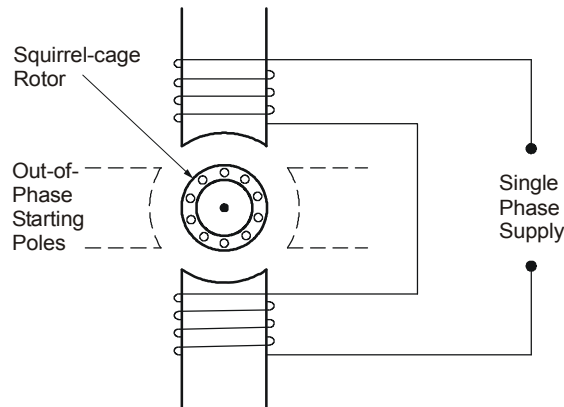
- a. Has a high starting current (reduced by a star-delta starter).
- b. Has a poor starting torque.
- c. Runs at almost synchronous speed, and the speed cannot easily be varied.
- d. Can be adapted for frequent reversing.

### Two-phase Induction Motors

44. A rotating magnetic field is also produced if two phases,  $90^\circ$  out of phase with each other, are used instead of a three-phase supply. The production of the rotating magnetic field in a two-phase induction motor employs similar methods to those described for the three-phase motor, and the motor principles are also similar to those of the three-phase motor. Two-phase motors are less efficient than three-phase motors, and therefore less widely used.

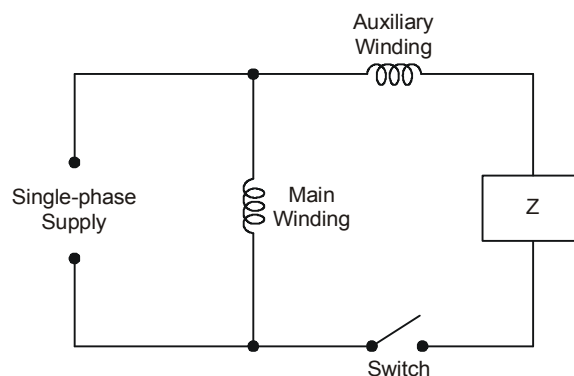
### Single-phase Induction Motors

45. Fig 17 represents a single-phase induction motor with one pair of stator poles and a squirrel-cage rotor. Such a motor is not capable of producing a rotating field in the manner previously described, and it is not self-starting.

**14-10 Fig 17 Single-phase Induction Motor**

46. The field produced by the single-phase winding alternates according to the frequency of the supply. As the field changes polarity every half-cycle, it induces currents in the rotor which try to turn it through  $180^\circ$  but, as the force is exerted on the rotor axis, there is no turning force on the rotor at rest. If the rotor is turned by external means to overcome its inertia, an impulse every half-cycle will keep it rotating. As the field is pulsating rather than rotating, single-phase motors are not as smooth running as two or three-phase motors.

47. The starting device often takes the form of an auxiliary stator winding spaced  $90^\circ$  from the main winding (Fig 17) and connected in series with an impedance to the main supply. This impedance is chosen to produce as great a phase displacement as possible between the currents in the main and auxiliary windings, so that the machine starts up virtually as a two-phase motor (Fig 18). A switch, usually operated by centrifugal action, cuts out the auxiliary winding when a fair speed has been attained, and the machine continues to run on the main stator winding. Single-phase induction motors (operating on 230 V 50 Hz supply and rated up to 5 hp) are used to provide a relatively constant-speed drive to DC generators.

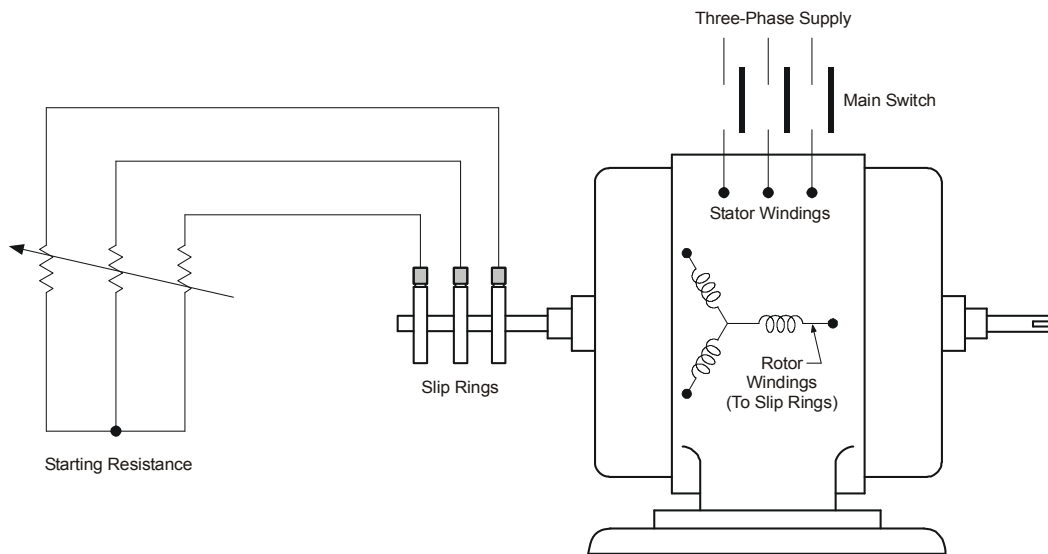
**14-10 Fig 18 Starting Device for a Single-phase Induction Motor**

### Wound Induction Motors

48. The squirrel-cage motor takes a large starting current, and has a poor starting torque, due to the low resistance rotor. These features are improved in the wound or slip-ring type of induction motor. The stators of wound induction motors are identical to those of squirrel-cage motors, but the rotor conductors are insulated and form a three-phase, star-connected winding, the three ends of which are connected to three insulated slip rings mounted on the motor shaft (Fig 19). When the motor is running normally, these slip rings are short-circuited to give a low resistance rotor equivalent to the squirrel cage. For starting, the slip rings are connected to a three-phase, star-connected resistance, as shown in Fig 19, and maximum resistance is inserted in the rotor circuit, improving the starting torque and giving a lower

starting current. The resistance is gradually cut out as the machine speeds up until, finally, the three slip rings are short-circuited and the motor runs as for a squirrel-cage machine.

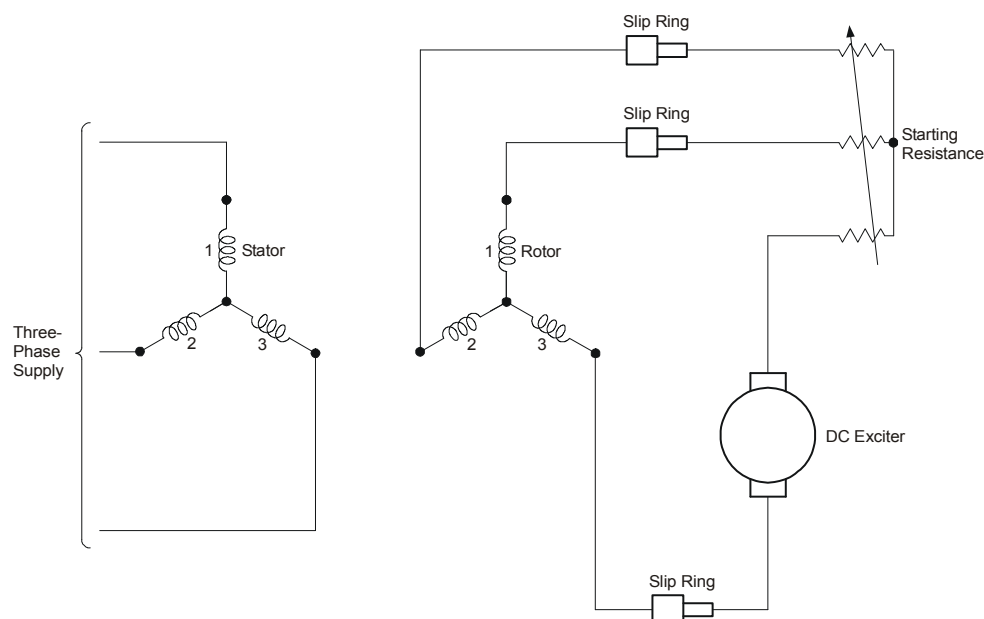
**14-10 Fig 19 Slip Ring Induction Motor**



### Synchronous Induction Motors

49. The features of constant speed and capacitor action of the synchronous motor are often combined with the self-starting feature of the induction motor by using a machine which is virtually a hybrid of the two types. One type of synchronous induction motor in wide use consists of a slip-ring induction motor coupled to a DC exciter; the rotor windings are connected, via the slip rings, to the exciter and a three-phase starting resistance, as shown in Fig 20. The exciter is driven by the motor, which is started up as a slip-ring induction motor. There is no appreciable DC from the exciter until the motor has attained some 90% of full speed. As synchronous speed is approached, the machine automatically pulls itself into synchronism, and continues to run as a synchronous motor, with DC applied to the rotor windings from the exciter. If the machine is overloaded, it pulls out of synchronism like an ordinary synchronous motor, but continues to run at reduced speed as an induction motor.

**14-10 Fig 20 Synchronous Induction Motor**



## Single-phase Commutator Motors

50. The induction and synchronous types of AC motor, although possessing many excellent features, are not variable-speed machines. Furthermore, the starting torque of a squirrel-cage motor is poor, it takes a large starting current, and the power factor is low. For these reasons, AC commutator motors have been developed, in an attempt to obtain improved speed control, starting torque, and power factor.

51. An ordinary DC series motor, when connected to an AC supply, will rotate and exert a unidirectional torque. The ordinary DC series motor would not, however, be capable of giving an efficient and satisfactory performance for the following reasons:

- a. Large eddy currents in the field magnets would cause undue heating.
- b. Each coil of the armature winding, when short-circuited by the brushes, would give rise to destructive sparking at the brushes.
- c. The power factor would be very low because of the highly inductive nature of the field and armature windings.

52. To overcome these drawbacks, a series motor designed for use on AC supply (or for use on either AC or DC, in which case it is known as a universal motor), is modified in the following manner:

- a. The entire magnetic circuit is laminated.
- b. The field winding is distributed in core slots, like the stator winding of an induction motor.
- c. The armature winding is sub-divided to a greater extent than in DC machines. This necessitates a relatively large number of commutator segments, and gives a large commutator for the size of the motor.
- d. The brushes are of high resistance carbon, and each brush is restricted in width so that it bridges only two commutator segments.
- e. Connection between armature coils and commutator segments is often made through resistors, instead of by direct soldering. This reduces the circulating current when a coil is short-circuited by a brush.

53. The main characteristics of the AC series motor are similar to those of DC series motors (Volume 14, Chapter 6), and the variable series resistor method for starting and speed control of the DC motor is used for the AC motor. AC commutator motors are not very efficient, and their use is confined to fractional horsepower motors. They are widely used in household appliances.

## ROTARY INVERTERS AND CONVERTERS

### Rotary Inverters

54. A rotary inverter combines the functions of a DC motor and an AC generator. It is similar in construction to a rotary transformer (Volume 14, Chapter 6), except that the generator windings are connected to slip rings. The function of the machine is to invert the DC input to an AC output.

### Rotary Converters

55. A rotary converter is a combination of an AC motor and a DC generator, and converts an AC input to a DC output.

## CHAPTER 11 - TRANSFORMERS

Contents	Page
Introduction .....	1
The Basic Voltage Transformer .....	1
Transformer Turns Ratio .....	2
Transformer Losses .....	2
Transformer Regulation .....	3
Three-phase Transformers .....	3
Auto Transformers .....	4
Transformer Matching .....	4
Current Transformers .....	4

### Table of Figures

14-11 Fig 1 A Typical Voltage Transformer .....	1
14-11 Fig 2 Voltages and Currents in a Transformer .....	2
14-11 Fig 3 The Development of a Three-phase Core .....	3
14-11 Fig 4 The Auto Transformer (Principle of Operation) .....	4
14-11 Fig 5 Transformer Matching .....	4
14-11 Fig 6 A Current Transformer .....	5

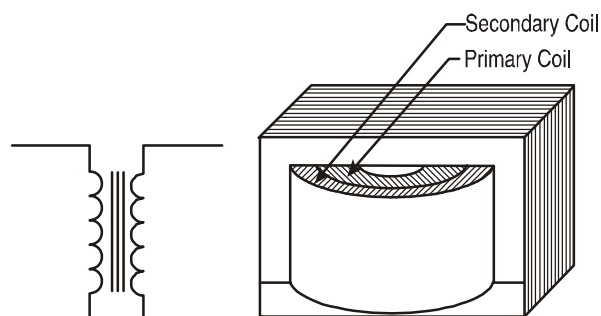
### Introduction

1. In the main, transformers are used as a means of increasing or decreasing AC voltages. They are devices with no moving parts and depend on mutual inductance between two electric circuits for their action. Transformers are not amplifiers or generators, as no external power supplies are used. The power out of the device is slightly less than the power put in to it.

### The Basic Voltage Transformer

2. The voltage transformer is a mutual inductor in which the coupling is arranged to be a maximum by concentrating the flux in an iron core, so that the flux through the input (primary) and output (secondary) windings is identical. Under ideal conditions, with the same flux in both primary and secondary windings, the transformer is said to have a coupling coefficient ( $k$ ) equal to unity. In reality, device losses have to be taken into account and unity is never achieved. Fig 1 shows the construction of a typical transformer along with its symbolic representation.

**14-11 Fig 1 A Typical Voltage Transformer**

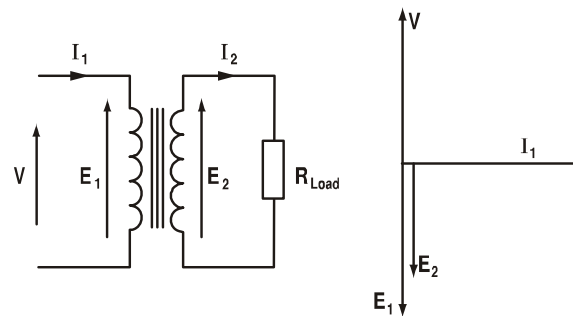


3. In the circuit shown in Fig 2, the applied AC voltage  $V$  causes a current  $I$  to flow in the primary which induces the same number of volts per turn in both primary and secondary windings of the transformer. The voltage induced in the primary,  $E_1$ , is a back emf and as such is in opposition to the applied voltage. If the primary has  $n_1$  turns in its windings and the secondary has  $n_2$ , then the voltage induced in the secondary,  $E_2$ , is given by:

$$E_2 = \frac{n_2}{n_1} \times E_1 \text{ (Back emf)}$$

The phasor diagram representation of this equation is also shown in Fig 2.

**14-11 Fig 2 Voltages and Currents in a Transformer**



### Transformer Turns Ratio

4. The number of turns in the secondary winding ( $n_2$ ) divided by the number of turns in the primary ( $n_1$ ) is known as the turns ratio of the transformer. The secondary voltage can be made many times larger or smaller depending on the turns ratio. Transformers are referred to as either step-up or step-down devices.

5. Some transformers are designed with tappings on their primary and secondary windings in order to accommodate a wide range of voltages. In these cases the transformer ratio is determined by the number of active windings in circuit for a given configuration.

### Transformer Losses

6. During the process of transformer action two phenomena occur within the metal core, namely hysteresis and eddy current effects. Both of these give rise to power losses within the device and are present whether a load is connected across the secondary or not. These losses are grouped under the heading of magnetization or iron losses, and, because they are associated with the input, they may be likened to the internal resistance losses of a battery or generator. These losses may be minimized as follows:

- a. Using magnetic materials with a high permeability and a small hysteresis loop area (see Volume 14, Chapter 2).
- b. Constructing a core made up of laminated sheets rather than a solid block. The laminations are insulated from each other, giving high resistance to the eddy currents tending to flow.

7. When current flows in the secondary circuit of a transformer, ie load connected, additional losses are introduced which are known as copper losses. Copper losses are caused by the resistance of the windings, and, because they are associated with the output, they may be likened to the load across a battery or generator.

8. For any transformer, maximum efficiency occurs, i.e. maximum transference of power, when the load losses (copper losses) equal the internal fixed losses (iron losses). This is very much in line with the maximum power transfer theorem stated in Volume 14, Chapter 1.

### Transformer Regulation

9. The rated voltage of a transformer is the voltage that appears across the secondary terminals on no-load, when the rated primary voltage is applied to the primary. However, the secondary voltage on-load will generally be less than this value; the drop depending on the load, the load power factor and the losses within the transformers. The value of the on-load voltage is particularly important because it is the value which applies in practice when the transformer is in operation with a load connected. The numerical value of the reduction is known as the inherent voltage regulation and is expressed as:

Secondary voltage, off-load ( $E_s$ ) – Secondary voltage, on-load ( $V_s$ ).

This is normally expressed as a percentage, or as a per-unit (pu) value.

$$\text{As a percentage} = \frac{E_s - V_s}{E_s} \times 100$$

$$\text{As a pu} = \frac{E_s - V_s}{E_s}$$

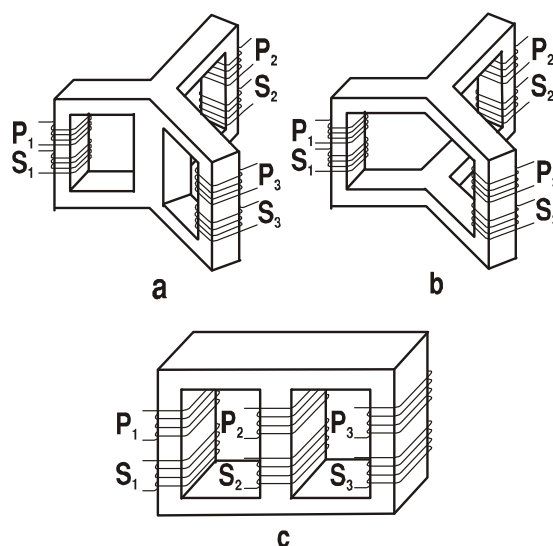
### Three-phase Transformers

10. In a three-phase AC system three separate transformers could be used in order to provide the necessary voltage transformation; however, it is more economical to use a single unit. It is usually more efficient, requires less space and weighs less than three transformers of the same capacity. Fig 3 shows the development of the three-phase core.

11. Fig 3a shows three single-phase transformers brought together to share a common central leg. However, the sum of the fluxes in the common leg will always be zero, and it is therefore not necessary. Fig 3b shows a combined core without the central leg, but this shape is difficult to make. The practical shape is shown at Fig 3c and is a distorted version of the previous form.

12. The primary and secondary coils of a three-phase transformer may be connected together in various star and delta configurations, depending on load and circuit requirements.

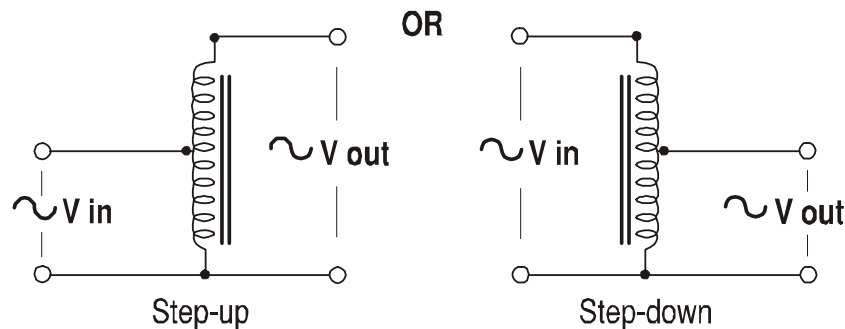
**14-11 Fig 3 The Development of a Three-phase Core**



## Auto Transformers

13. Any tapped inductance can be used as a transformer, and this is illustrated in Fig 4. This device is called an auto transformer and the action is similar to that of the conventional type. The main difference is that part of the inductance is common to both primary and secondary circuits and as such provides direct coupling between them. Primary and secondary currents are essentially in antiphase and equal in value when the turns ratio is zero. The inductance then carries zero current. The auto transformer is particularly efficient for small transformer ratios.

14-11 Fig 4 The Auto Transformer (Principle of Operation)



## Transformer Matching

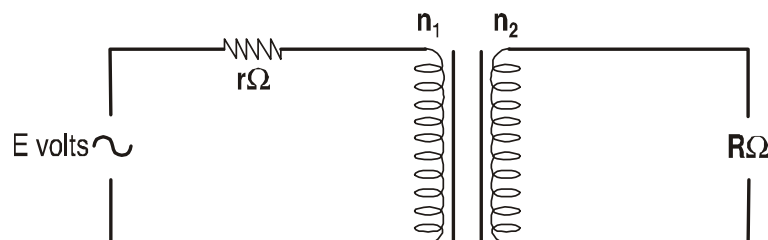
14. The maximum power transfer theorem established that maximum transference of power takes place when the resistance of the load is equal to the internal resistance of the supply source. Under these conditions the power source and load are said to be matched.

15. Unfortunately, in practice, circuit loads very rarely equal the internal resistance of the supply and transformers are used with suitable turns ratios to provide the required matching. This is shown in

Fig 5. A turns ratio  $\left( t = \frac{n_2}{n_1} \right)$  is chosen so that:

$$r = \frac{R}{t^2} \text{ or } t = \sqrt{\frac{R}{r}}$$

14-11 Fig 5 Transformer Matching

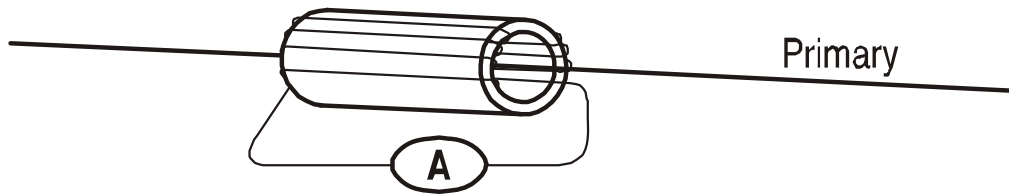


## Current Transformers

16. Current transformers, sometimes referred to as instrument transformers, utilize the current or voltage transformation properties of the normal transformer. The main difference in operation is that the primary winding forms part of the circuit under test (see Fig 6).



**14-11 Fig 6 A Current Transformer**



17. Current transformers are used for:
- a. Measuring current or power in high voltage systems.
  - b. Measuring high current with low range instruments.
  - c. Operating relays and other control devices according to the current in a system.

## CHAPTER 12 - FUNDAMENTAL ELECTRONIC COMPONENTS

Contents	Page
Introduction .....	2
SEMICONDUCTORS .....	2
Introduction .....	2
Doped Semiconductors .....	2
N-Type and P-Type Semiconductors .....	2
P-N Junctions .....	3
Properties of P-N Junctions .....	4
P-N Junction Diode Characteristics .....	5
Junction Diode as a Switch .....	5
TRANSISTORS .....	6
Introduction .....	6
Transistor Junctions and Bias Voltages .....	6
Current Flow in Transistors .....	7
Advantages of Transistors .....	8
FIELD EFFECT TRANSISTORS .....	8
Introduction .....	8
The Junction Gate FET .....	8
The Insulated Gate FET .....	10
INTEGRATED CIRCUITS .....	10
Introduction .....	10
Composition .....	10
Applications .....	11
Development .....	11
CATHODE RAY TUBES .....	11
Introduction .....	11
The Electrostatic CRT (ESCRT) .....	12
The Electromagnetic CRT (EMCRT) .....	14
The ESCRT and EMCRT Combined .....	14
CRT Displays .....	15
<b>Table of Figures</b>	
14-12 Fig 1 Flow of Electrons and Holes .....	3
14-12 Fig 2 Direction of Flow .....	3
14-12 Fig 3 Effect of Bias in P-N Junction Diode .....	4
14-12 Fig 4 Typical Diode Characteristics .....	5
14-12 Fig 5 Two Types of Transistor .....	6
14-12 Fig 6 Transistor Junction .....	6
14-12 Fig 7 Current in NPN Transistor .....	7
14-12 Fig 8 Basic Construction of the FET .....	8
14-12 Fig 9 How the FET Works .....	9
14-12 Fig 10 MOSFET .....	10
14-12 Fig 11 Piece of an Integrated Circuit .....	11
14-12 Fig 12 The Electrostatic CRT (ESCRT) .....	12
14-12 Fig 13 Post-deflection Acceleration .....	13
14-12 Fig 14 The Electromagnetic CRT (EMCRT) .....	14
14-12 Fig 15 A-scope Waveform .....	15

## Introduction

1. Electronics is the science and technology of controlling suitably modified electron flow so as to convey information. A good deal of electronic equipment looks complicated and complex on first examination, but it has to be realized that all electronic circuits are made up of relatively simple basic units each performing a specific function. The way in which these units, or 'building blocks', are assembled governs the function of the complete system.
2. The whole of modern electronic circuitry is built on a foundation of resistors, capacitors, inductors, and semiconductor devices. From these simple components, intermediate units are assembled, such as:
  - a. Amplifiers, to increase the size of signals.
  - b. Oscillators, to generate desired waveforms and signal frequencies.
  - c. Mixers, to combine signals of different frequencies.
  - d. Power supplies, to energize the whole system.
  - e. Modulators, to superimpose information on radio waves.
  - f. Demodulators (detection), to remove information from radio waves.
  - g. Transducers, to convert a physical quantity (force, light, sound, etc) to an electrical signal, and vice versa.
  - h. Logic gates, to perform logic functions.
  - i. Shift registers, for use in microprocessors.
3. Electronic components fall into two main categories, namely active and passive devices. Resistors, capacitors, and inductors are regarded as being 'passive', while transistors and allied semiconductor devices are regarded as 'active', because they modify the power supplied to them. This chapter is devoted to the subject of active devices; passive devices are covered in detail in Volume 14, Chapter 1.

## SEMICONDUCTORS

### Introduction

4. Most materials used in electricity and electronics are either conductors or insulators. However, a few materials do not fall into either of these categories because conduction through them is too small for them to be classed as conductors, and too large for them to be insulators. Such materials are known as semiconductors. Germanium and silicon are the two most common semiconductors.

### Doped Semiconductors

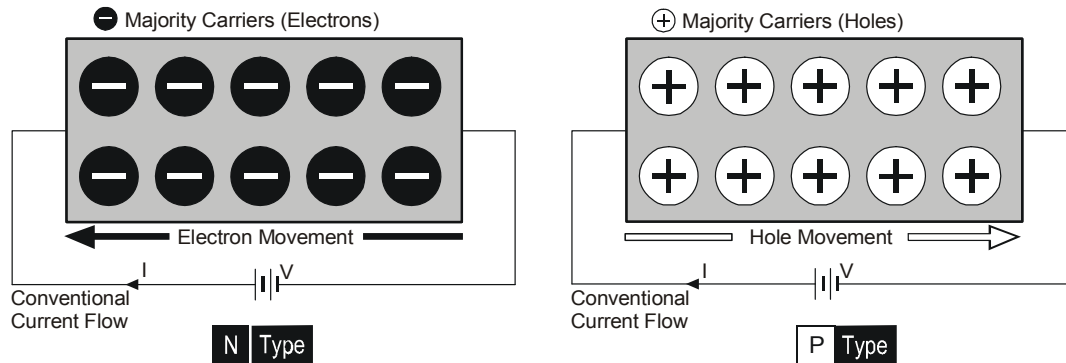
5. At ordinary room temperatures, a pure semiconductor has few free electrons. However, if very small quantities of another selected element, such as indium or arsenic, are combined with the pure semiconductor, conductivity is increased. This process of adding 'impurities' to the pure germanium or silicon is known as 'doping'. The resulting conductivity of the semiconductor depends upon the amount of doping, and can be strictly controlled.

### N-Type and P-Type Semiconductors

6. Doped semiconductors can be one of two types (n-type or p-type) depending upon the element that is added during doping. For an n-type semiconductor, the result of doping is to produce free electrons which, as negative charges, become available as n-type 'charge carriers' (n for negative). Conversely, for a p-type semiconductor, doping produces positive (or p-type) charge carriers known

as 'holes'. A hole has a positive charge, equal and opposite to that of an electron. If a voltage is applied to an n-type semiconductor, the resulting current is due mainly to a movement of electrons; in a p-type, it is due mainly to a movement of holes (Fig 1). The moving charges in each case are known as 'majority carriers'.

14-12 Fig 1 Flow of Electrons and Holes



**Note:** While hole movement is in the **same** direction as conventional current flow, electron movement is in the **opposite** direction.

7. Although the majority of charge in an n-type is carried by electrons, there is a small flow of charge due to hole movement, known as 'minority carriers'. Similarly, minority carriers are electrons in p-type material where the majority carriers are holes. Minority carriers can be ignored when considering the basic operation of semiconductor devices, but they are very important when determining breakdown voltages, heating, and the efficiency of devices.

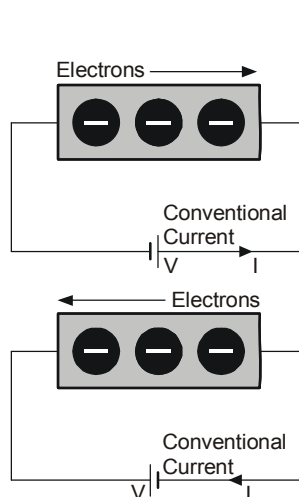
### P-N Junctions

8. A single piece of n-type or p-type semiconductor conducts readily in either direction; reversing the applied voltage merely reverses the direction of current (Fig 2a).

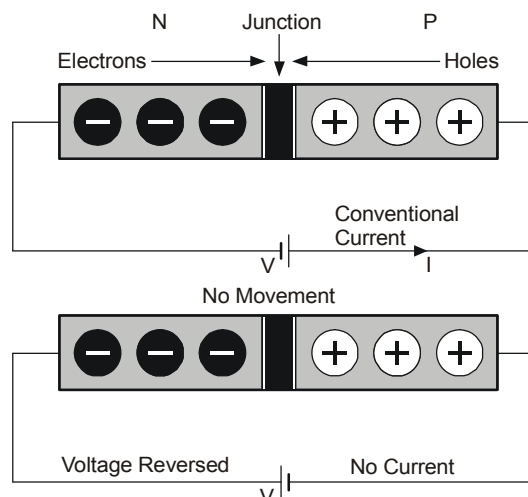
9. In a continuous piece of pure semiconductor that has been doped in such a way that one end of the semiconductor has n-type properties and the other end p-type, conduction takes place easily in only one direction (Fig 2b). This is the basis of the p-n junction diode. Note that it is not sufficient merely to join a piece of n-type semiconductor to a piece of p-type, the junction must occur in a continuous piece of single-crystal semiconductor.

14-12 Fig 2 Direction of Flow

#### a N Type (P Type Similar)



#### b P-N Junction



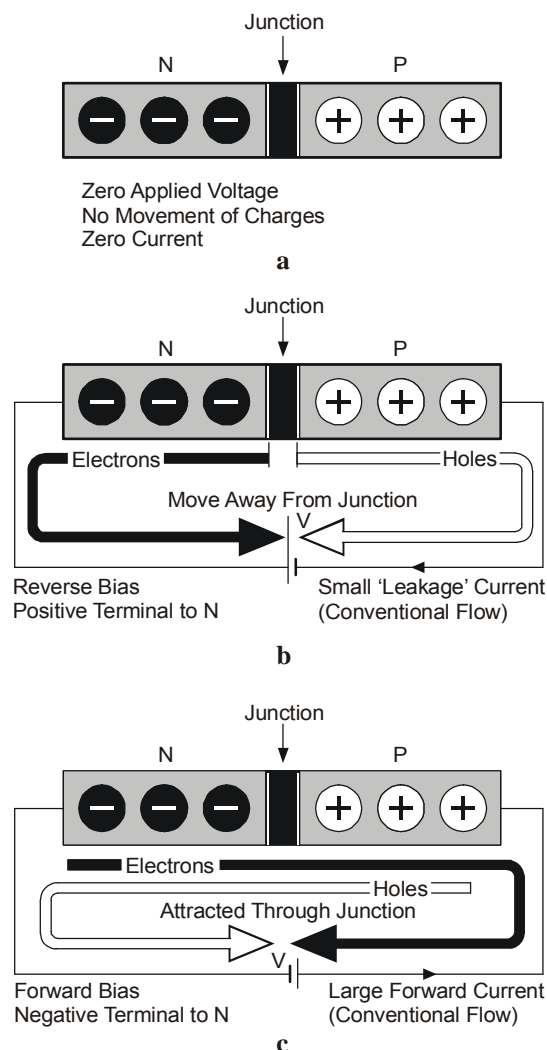
## Properties of P-N Junctions

10. When a p-n junction is formed, the distribution of charges in the n and p-type materials is such that there is no net flow of current across the junction. Thus, under the condition shown in Fig 3a, with no external voltage applied, there is no current flow.

11. If a battery is connected across the semiconductor slice, as shown in Fig 3b, the positive terminal attracts electrons in the n-type region away from the junction, and the negative terminal of the supply has the same effect on the holes. There is now even less possibility of current flow through the p-n junction. A junction diode operating under these conditions is 'reverse biased'. In practice, in a reverse biased diode, a very small reverse current flows (measured in micro amps) and this is called the leakage current; due to minority charge carriers.

12. If the external battery is reversed, electrons are attracted from the n-type region through the junction to the positive terminal of the battery, and holes are attracted the other way. The two movements combine to give a large forward current through the junction in the direction shown. A junction diode operating under these conditions is 'forward biased', as shown in Fig 3c. A large current (normally measured in milliamps) flows for a small forward bias voltage; this indicates that the diode has a low resistance in the forward direction.

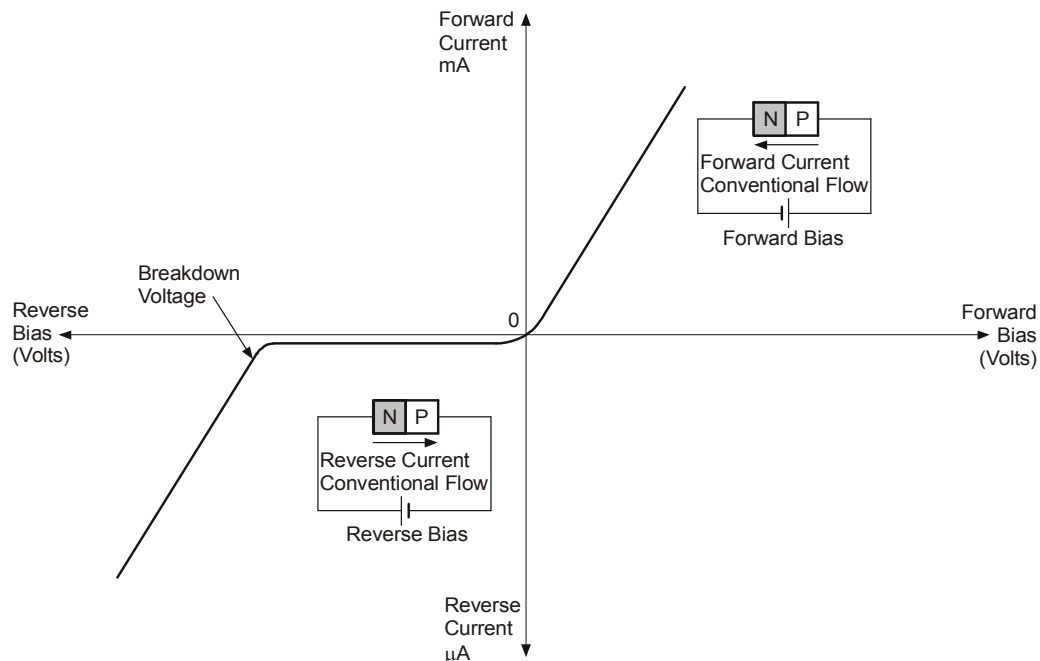
**14-12 Fig 3 Effect of Bias in P-N Junction Diode**



## P-N Junction Diode Characteristics

13. By measuring the current flowing in the external circuit for various values of reverse and forward bias voltages applied to a diode, we can obtain a series of readings relating the current and the voltage. From these readings, a 'characteristic' of the diode may be plotted showing how the diode current varies with the voltage applied to it (Fig 4).

**14-12 Fig 4 Typical Diode Characteristics**



**Note:** Because forward current scale is in mA and the reverse current scale is in  $\mu\text{A}$ , reverse current is scaled up 1,000 times. Without this scale magnification, the reverse current would appear to be zero up to the breakdown point.

14. Fig 4 shows that current flows more easily in the forward direction than in the reverse direction. After the initial 'bend' around zero volts, the current in the forward direction is proportional to the forward bias voltage and the graph becomes approximately a straight line. When the diode is reverse biased, only a very small reverse (leakage) current flows until a certain 'breakdown' point is reached. At this point, the reverse current rises to a high value, often sufficient to destroy the diode. The reverse voltage at which the diode breaks down depends upon the construction of the device, but a diode must not be used in circuits where the reverse voltage that can be applied to it exceeds its breakdown voltage.

## Junction Diode as a Switch

15. When a p-n junction is forward biased, its resistance is very low, often a fraction of an Ohm. When reverse biased, the very small leakage current, even at high reverse voltages, means that the resistance in this direction is very high, of the order of megohms. Thus, by applying suitable forward and reverse voltages to a diode, it may be used in the same way as a switch: ON when forward biased; OFF when reverse biased. Diode switches have many uses in electronics. The switching is faster than that of a mechanical switch, the diode may be very small, and there are no moving parts to wear out.

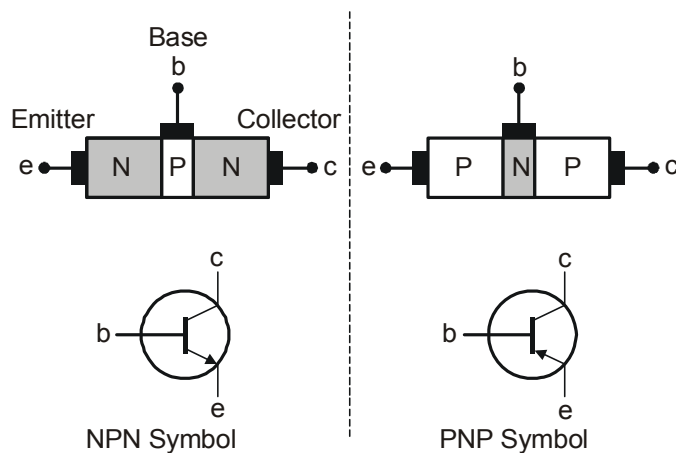
16. For rapid switching (of the order of picoseconds) at high powers (up to 10 kW peak), a PIN diode is often used. This has an undoped (intrinsic) layer sandwiched between the p and n layers. This layer alters the diode characteristics to give the high speed, high power requirements.

## TRANSISTORS

### Introduction

17. A semiconductor diode is a two-electrode device, one end having p-type properties and the other end n-type. A transistor is a three-electrode device and has three layers of semiconductor, alternately p-type and n-type. The three layers merge into one another to form a sandwich (Fig 5). One of the outer layers is known as the 'emitter', the other is the 'collector', and the thin centre layer is known as the 'base'.

**14-12 Fig 5 Two Types of Transistor**



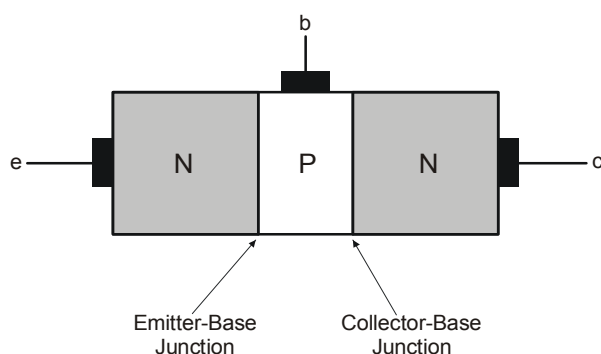
18. There are p-n-p or n-p-n transistors. In an n-p-n transistor, the emitter and collector are n-type semiconductors and the base is p-type. In a p-n-p transistor, the semi-conductor materials are interchanged. The circuit symbols are also shown in Fig 5. The arrow on the emitter points in the direction of conventional current flow; for an n-p-n transistor, this is into the emitter, whilst in a p-n-p type it is away from the emitter.

19. A transistor has many functions: it may be used as a switch, as a control device, or as the active device in amplifier and oscillator circuits. It has the advantages of being very small and light; but, even more important, it is reliable, it requires very little power for its operation and, if operated correctly, has a long life.

### Transistor Junctions and Bias Voltages

20. In a semiconductor junction diode, there is one p-n junction. In a transistor, we have two: one between the emitter and the base, the other between the collector and the base (Fig 6).

**14-12 Fig 6 Transistor Junction**



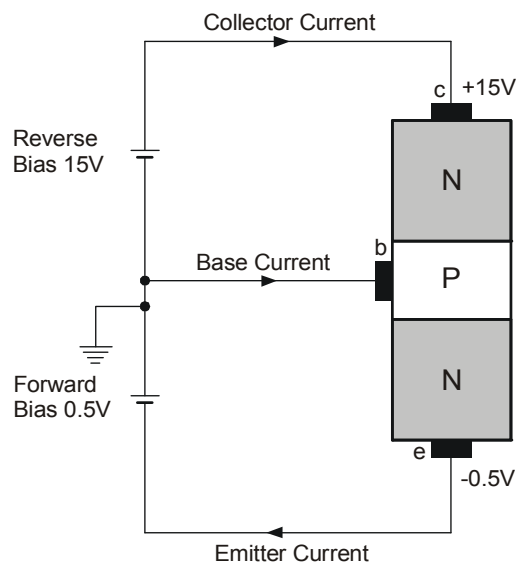
21. For a transistor to operate, it must have suitable voltages applied across each junction, and these voltages must be of the correct polarity; in other words, the transistor has to be properly biased. A p-n junction is forward biased when the p-type semiconductor is connected to the positive terminal of the supply; under this condition, the resistance of the junction is low and current flows easily. A junction is reverse biased when the p-type semiconductor is connected to the negative terminal of the supply; the resistance of this junction is high and the only current flowing is that due to leakage.

22. Under operating conditions, the emitter-base junction in a transistor is forward biased and the collector-base junction is reverse-biased. This applies whether the transistor is n-p-n or p-n-p. Because of the forward and reverse bias, the emitter-base junction has a low resistance and the collector-base junction a high resistance.

### Current Flow in Transistors

23. Transistor action relies upon a controllable current flowing between collector and emitter. Consider the n-p-n configuration shown in Fig 7. Electrons are flowing from the emitter to the base across the forward-bias junction. These electrons would normally flow out of the base, and could not go on to the collector because the base-collector junction is reverse biased. However, if the base junction is very thin, there are more electrons arriving from the emitter than there are positive holes in the p-type material. The base-collector junction then effectively becomes forward biased, so most of the free electrons in the base can continue their journey to the collector.

**14-12 Fig 7 Current in NPN Transistor**



24. There is a large current between collector and emitter and a small current into the base. Controlling the current which can enter the base, controls the junction biasing and the current which flows between the collector and the emitter, ie a small base current controls a large collector-emitter current.

25. Similar remarks apply to a p-n-p transistor, except that the batteries must be reversed to provide the correct bias. The forward-biased p-type emitter injects holes into the n-type base, with most of the holes moving through the base region to be collected by the collector.



## Advantages of Transistors

26. Transistors and their allied semiconductor devices have now replaced thermionic valves in almost every type of electronic equipment. Their success has been due to the great advantages they offer over valves, that is:

- a. Much smaller size.
- b. They do not require heating.
- c. The ability to withstand greater levels of vibration and shock.
- d. No warming up period.
- e. Supply voltages are much lower.
- f. Longer life.
- g. They are better suited for use in remote locations, such as space satellites and booster/repeater amplifiers.

27. The early problems associated with limited power output, noise, frequency response, and operating temperature are all being overcome as research intensifies. Nowadays there are very few applications in which a thermionic device is superior to a transistor.

## FIELD EFFECT TRANSISTORS

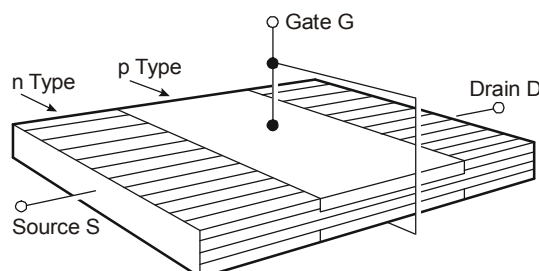
### Introduction

28. The type of transistor so far discussed is often called the 'bipolar' transistor because there is a simultaneous flow of both main and leakage currents. Another type of transistor which has been introduced has certain advantages and disadvantages when compared with the bipolar type. It is known as the field effect transistor (FET). There are basically two types of field effect transistor, the junction-gate (or 'jugfet') and the insulated gate FET (or 'igfet').

### The Junction Gate FET

29. The schematic outline of a basic FET is shown in Fig 8. It consists of a slice of n-type silicon into which two p-type regions are inserted opposite each other. The p-type regions are joined together externally and form a control electrode known as the gate (G). One end of the slice is referred to as the source (S) and the other end is the drain (D). The source is comparable to the emitter in a bipolar transistor, the gate to the base, and the drain to the collector.

**14-12 Fig 8 Basic Construction of the FET**

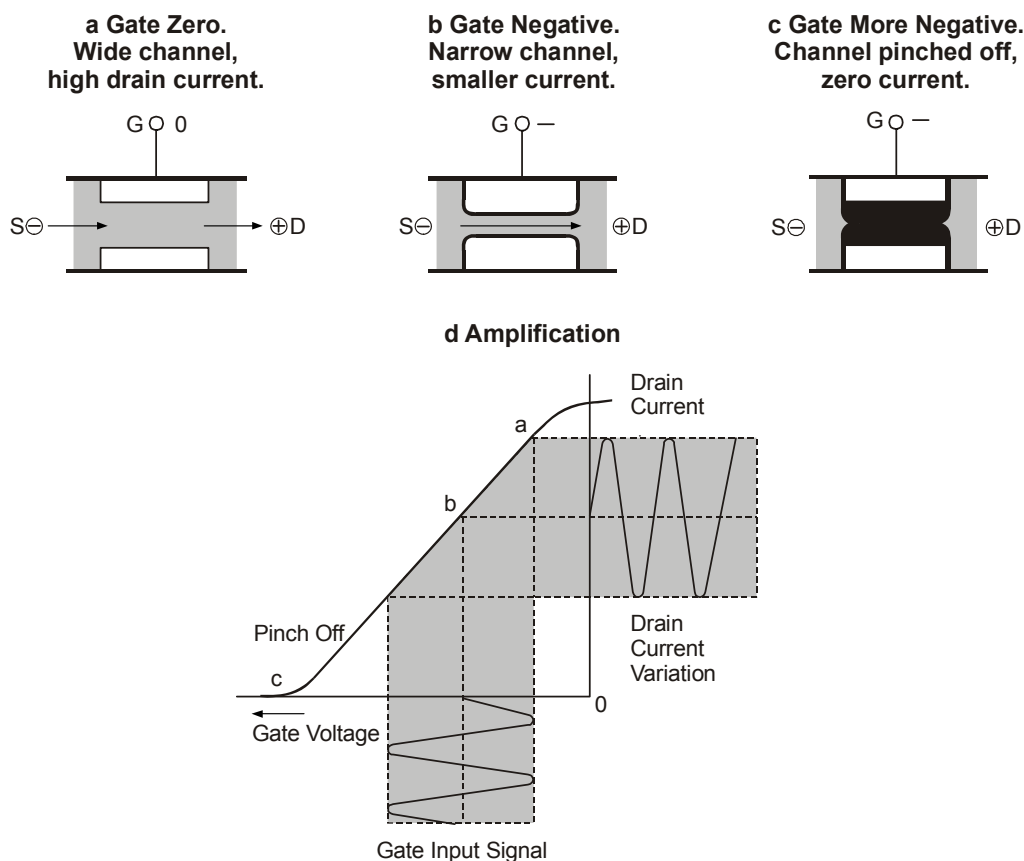


30. Fig 9 indicates how the FET works. It normally has a conducting n-type channel between S and D. Therefore, if D is made positive relative to S, charge carriers (electrons in this case) flow from S to D, channelling between the p-type gate regions. If G is made negative to S, it effectively repels the moving electrons and 'squashes' them into a narrower funnel. In other words, the conducting channel becomes narrower and the drain current decreases. By making G more negative, a point is reached where the channel is 'pinched off' altogether and the drain current falls to zero.

31. If a signal voltage is applied to G, the drain current is caused to vary in sympathy. By suitable choice of load resistor in the drain circuit, the output voltage variations produced by the drain current can be much larger than the applied signal voltage. Amplification may therefore be obtained. The characteristics of a FET differ from bipolar transistors in three ways:

- a. The bipolar transistor is a current-operated device, ie it depends for its operation on the current applied as input to the emitter-based circuit. For a current to flow for a small signal voltage, the input resistance of the device must be low. In a FET, there is no current in the gate electrode. Thus, the FET is a voltage operated device, and has a very high input impedance, which is a requirement in many circuits.
- b. The FET has practically no leakage current.
- c. Transit time effects are negligible, so that the FET has a high cut-off frequency.

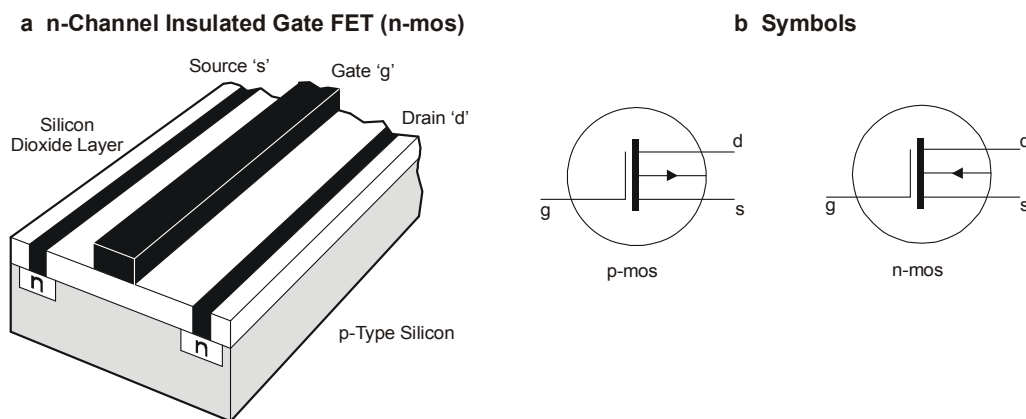
**14-12 Fig 9 How the FET Works**



## The Insulated Gate FET

32. Junction FETs are now only manufactured for replacement parts; all new FETs are of the Insulated Gate family. The basic construction of an insulated gate FET is shown in Fig 10. In this type of FET, the gate is insulated from the channel and, in the example shown, the source and drain connections are to two n-regions contained within a p-region. The gate is a metallic layer separated from the p-region by an insulating layer of silicon dioxide. The sequence of materials - metal, oxide, semiconductor - has given the device its alternative name of 'mosfet'. Although this is an n-channel device, there is, in fact, no n-layer between the source and the drain until the gate is biased positively with respect to the p-region (known as the 'base' or 'substrate'). Such a potential attracts free electrons from the n and p-regions to the upper surface of the p-region, where they form an n-channel bridging the two n-regions and so providing a conducting path between source and drain. It is significant that in the absence of a gate bias, no drain current can flow; a positive gate bias is necessary to give a drain current.

14-12 Fig 10 MOSFET



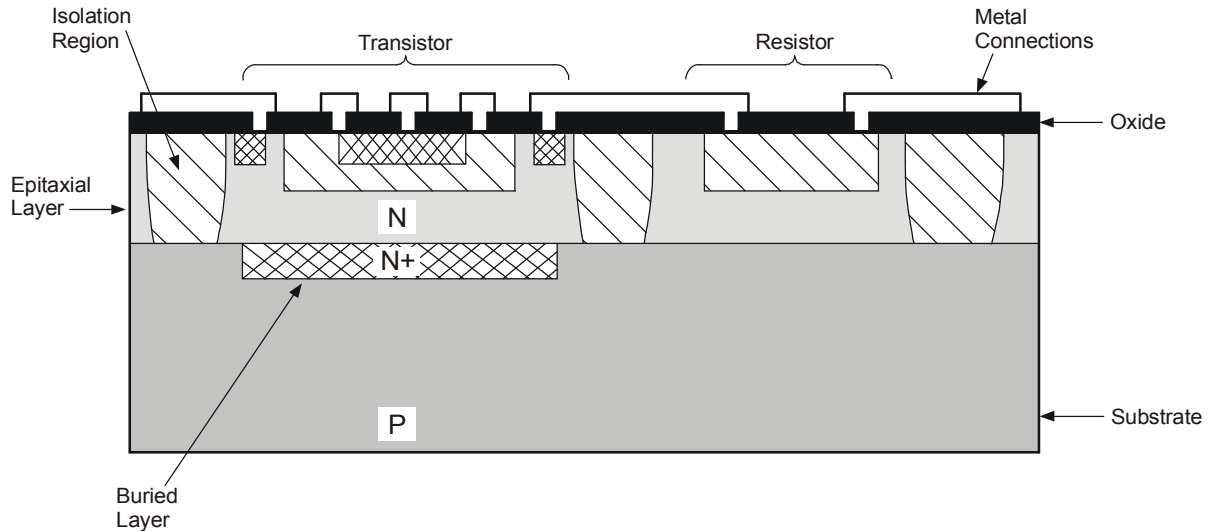
## INTEGRATED CIRCUITS

### Introduction

33. The twin requirements for small size and reliability, particularly in military systems, have led to many different approaches to the miniaturization of electronic circuits. Perhaps the most important of these is the monolithic silicon integrated circuit.

### Composition

34. Silicon integrated circuits are manufactured by a diffusion process developed from the planar technique, used for making transistors. In this process, hundreds, or even thousands, of transistors and other components can be diffused into a single silicon slice, which is then separated into individual circuit wafers with dimensions typically  $1.5 \text{ mm} \times 1.5 \text{ mm} \times 0.3 \text{ mm}$ . These components are connected together in the desired circuit configuration by a network of metal interconnections on the surface of a wafer. Fig 11 represents a simplified cross-section of a piece of such an integrated circuit containing a resistor and a transistor. Diodes can also be incorporated into integrated circuits, and small-value capacitors can be provided by using reverse-biased diodes which exhibit capacitance of a few pF. Larger value capacitors and inductors must be added externally to the diffused integrated circuits. A far greater degree of miniaturization can be achieved using MOS techniques; because of their high input resistance and simple construction, thousands of these devices may be diffused on a single silicon slice.

**14-12 Fig 11 Piece of an Integrated Circuit**

### Applications

35. The range of applications for integrated circuits is almost limitless; they can be used to provide all the basic computer circuits, such as NAND and NOR gates, and binary counters. Indeed, it is the widespread use of integrated circuits which has made possible high capacity, small size computers. There are also integrated circuits for single and multiple linear amplifier stages, pre-amplifying circuits and operational amplifiers.

### Development

36. A development in the field of integrated circuits is the concept of forming the system wiring between integrated circuits while they are still on the whole silicon slice, so that a single slice will form a complete electronic sub-system. This technique is generally called large-scale integration (LSI). It is important because one of the major causes of failure in electronic equipments is failure in the interconnecting wiring. The integrated circuit, with its system of interconnections within each circuit (formed as part of the slice fabrication process), gave a significant improvement in reliability. It is a logical step, in order to obtain still higher reliability, to produce the interconnections between circuits while they are still on the slice. This concept has been extended to cover very-large-scale integration (VLSI) and extra-large-scale integration (ELSI).

## CATHODE RAY TUBES

### Introduction

37. The cathode ray tube (CRT) is an electron tube in which electrons emitted from a cathode are formed into a narrow beam, accelerated to high speed, and directed to a screen coated with a fluorescent material. This material glows at the point of impact, producing a visible dot. A changing field, either electrostatic or electromagnetic, between the source of electrons and the screen causes the dot of light to move in accordance with the field variations.

38. There are two basic types of CRT; the electrostatic CRT and the electromagnetic CRT. In the electrostatic CRT, focusing and deflection are done with electronic fields, while in the electromagnetic CRT, magnetic fields do these jobs.

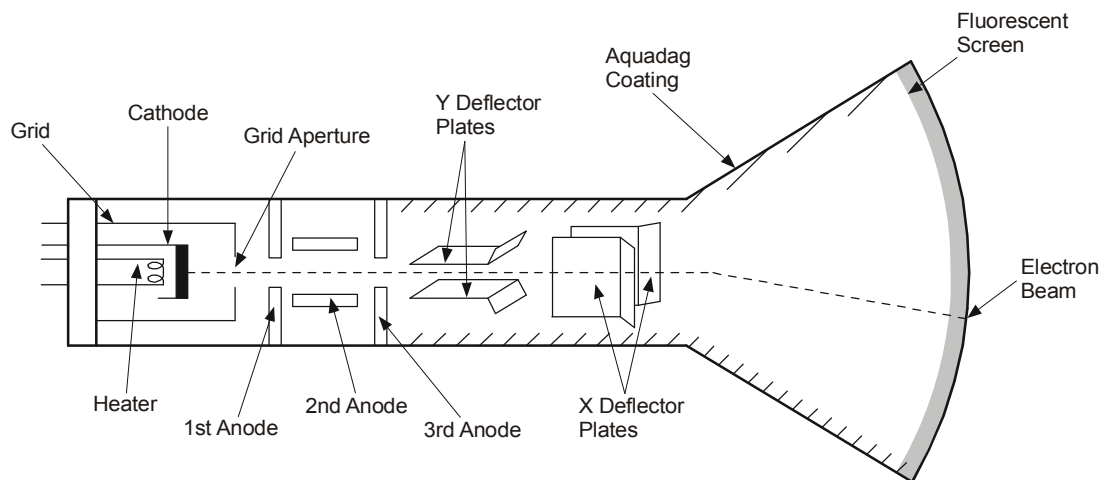
## The Electrostatic CRT (ESCRT)

39. Fig 12 shows the main parts of an electrostatic CRT. These parts can be divided into three groups:

- The electron gun, which produces a stream of fast-moving electrons and focuses them into a narrow beam. Sometimes the electron gun refers only to the cathode and grid; the anodes are then called the focusing system.
- The deflecting plates, which enable the beam of electrons to be moved up and down and from side to side.
- The fluorescent screen, which shows the movement of the beam by producing a spot of light.

All of the electrodes are enclosed in an evacuated glass envelope.

**14-12 Fig 12 The Electrostatic CRT (ESCRT)**



40. The cathode is a small tube, coated at the end with an oxide which emits electrons when heated. The grid is a hollow cylinder surrounding the cathode, with a central hole through which the electrons pass. The grid is made negative with respect to the cathode and, by varying this voltage, the number of electrons in the beam is varied, thus controlling the brilliance of the spot of light on the CRT screen. The brilliance control alters the voltage on the grid of the CRT. If the grid is made sufficiently negative, the electron beam will be completely cut off, and the spot on the screen will be blanked out.

41. The first and third anodes are circular plates, with holes through their centres. They are held at a high positive voltage relative to the cathode (several hundred, or even thousand, volts), and so they accelerate the electrons to a high speed. The third anode voltage is higher than the first anode voltage.

42. The second anode is a hollow cylinder mounted between the first and third anodes. Its purpose is to focus the electrons into a narrow beam and, for this reason, it is made negative with respect to the other anodes; this voltage can be varied to adjust the focusing of the beam. Hence, the focus control varies the voltage on the second anode.

43. In practice, the third anode is earthed and the other electrodes are made negative with respect to it. Typical EHT (extra high tension) voltages used in CRTs are: cathode  $-4$  kV; grid  $-4.02$  kV (variable); first anode  $-2$  kV; second anode  $-3$  kV (variable). Third anode and screen OV EHT voltages are needed in order to give the electrons enough speed to produce light on the fluorescent screen.

44. Two sets of deflecting plates are mounted after the third anode, as shown in Fig 12. By applying varying voltages to these plates, the focused beam can be swung in any direction. The plates nearer the third anode (the Y plates) are used to move the beam vertically, and those nearer the screen (the X plates) deflect the beam horizontally. The plates are often flared at the ends to provide the required amount of deflection without fouling the beam. When a voltage is applied to the plates, the electron beam will be attracted or repelled, depending on the polarity.

45. For horizontal deflection, a sawtooth voltage is applied to the X plates. This voltage increases from minimum to maximum values at a uniform rate (the sweep) and then returns rapidly to minimum (the fly-back). The waveform is repetitive and causes the beam to move from the left-hand to the right-hand side of the screen and then to return quickly to the left-hand side to start another sweep. The fly-back may be 'blanked out' by applying either a negative pulse to the grid or a positive pulse to the cathode during this period.

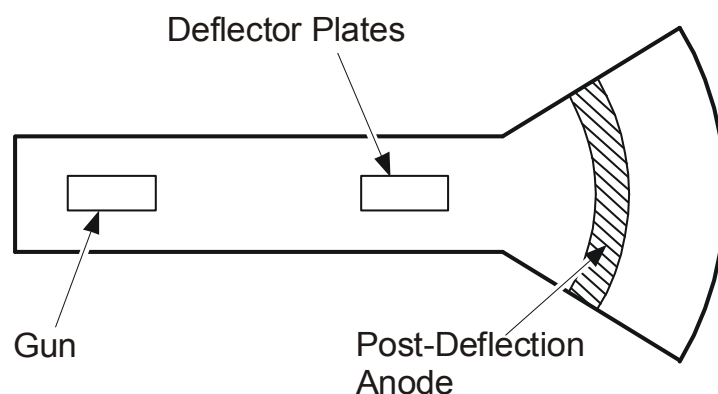
46. Thus, with this waveform applied to the X plates, a horizontal line is produced on the screen. When a voltage pulse is applied to the Y plates, the beam is vertically deflected for the duration of the pulse and a 'blip' appears on the screen.

47. The fluorescent screen is a chemical coating on the inside end of the glass tube. When fast-moving electrons hit the screen, they cause it to glow with a colour which depends upon the type of chemical used. The spot of light remains for a time after the electron beam has moved away; this effect is after-glow and enables a complete steady picture to be seen. In sonic types of radar display, the time base trace rotates slowly round the centre of the CRT screen and a chemical with a long after-glow is used to retain the picture.

48. When electrons hit the screen, secondary electrons are emitted. These are conducted via a powdered graphite coating, called the 'aquadag', to the third anode (earth). The coating prevents the screen becoming negatively charged.

49. In some types of CRT, an extra anode is used. It is made of a conducting ring of powdered graphite held at about twice the voltage of the third anode with respect to the cathode. This enables small input signal voltages to produce large displays on the screen, ie the deflection sensitivity of the tube is increased (Fig 13).

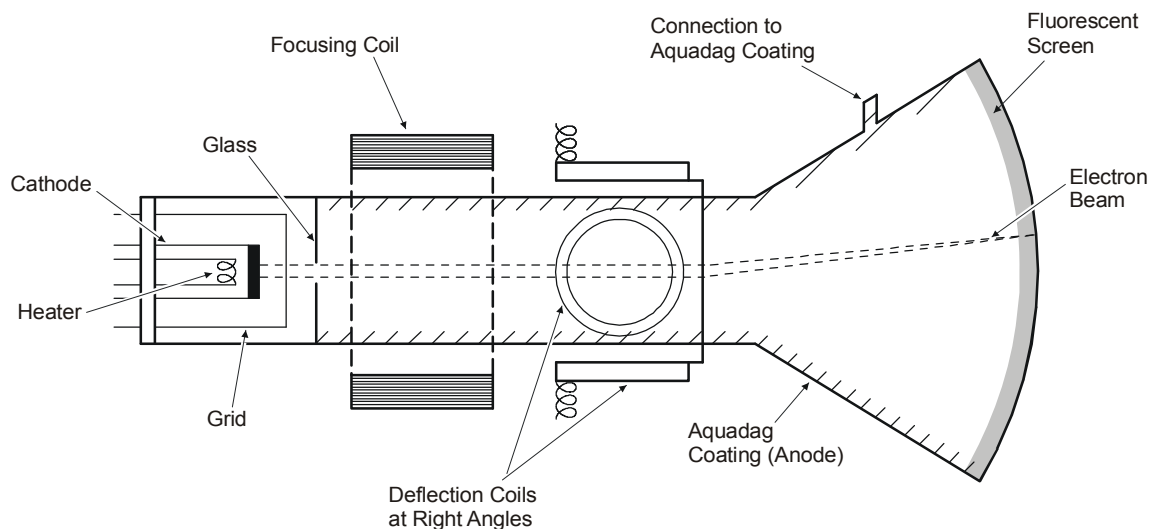
**14-12 Fig 13 Post-deflection Acceleration**



## The Electromagnetic CRT (EMCRT)

50. The construction of an electromagnetic CRT, employing magnetic focusing and deflections, is shown in Fig 14. The heater, cathode, and grid are the same as for an electrostatic CRT, but there is usually only one anode. This may be the aquadag coating, connected to a suitable voltage. Deflecting and focusing currents are passed through coils mounted on the outside of the neck of the tube. The electron beam is acted upon by a sideways force as it passes through the magnetic field around the coil, in much the same way as a current-carrying conductor has a force exerted on it in an electric motor.

**14-12 Fig 14 The Electromagnetic CRT (EMCRT)**



51. Focusing is done by passing DC through a specially shaped coil. The position of the coil, and the amount of current, control the focusing of the beam. The focus control is a rheostat, which varies the current through the coil.

52. Deflection is produced by two pairs of coils, mounted at  $90^\circ$  to each other, round the neck of the tube. In Fig 14, only one pair is shown for clarity. The amount of deflection depends on the strength of current in the coils. If the current is reversed, the direction of deflection reverses. To produce a horizontal time base, a sawtooth waveform of current must be passed through the horizontal deflection coils. Vertical deflection can be obtained by passing a signal current through the vertical deflection coils, but it is more usual to show signals on an electromagnetic CRT by using intensity modulation. In this case, a positive-going signal voltage is applied to the grid of the CRT. The number of electrons in the beam increases for the duration of the signal, and a bright spot appears on the time base.

## The ESCRT and EMCRT Combined

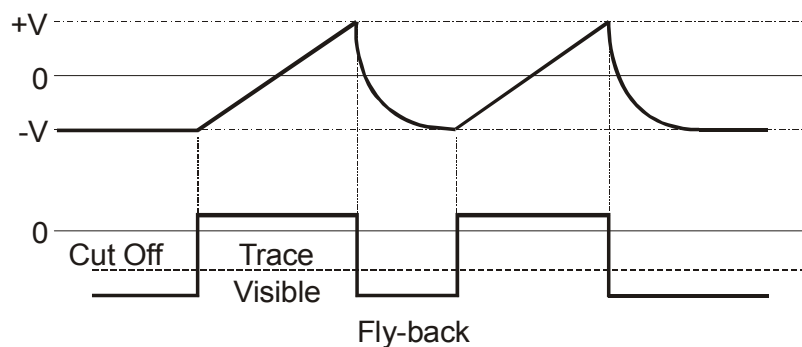
53. Combinations of the two CRT systems may be used, the most popular being that using electrostatic focusing and electromagnetic deflection. There are various advantages and disadvantages of the two types, and, until recently, the greatest advantage of the electromagnetic tube has been its higher resolution, ie ability to reduce the spot size to very small proportions. However, recent research has improved the resolution of the electrostatic CRT, although it has increased the complexity of the tube and necessitated the use of low-power magnetic correcting coils. Present-day electromagnetic tubes have spot sizes of about 0.25 mm.

## CRT Displays

54. **Time Base Production.** All CRT displays need some form of time base; a large number of displays use the A-scope time base. With this type, the spot moves linearly across the face of the screen, usually from left to right, then returns rapidly to its starting point. During the rapid return, known as 'fly-back', the spot of light is normally extinguished by application of a suitable blackout waveform to the grid of the tube. The ESCRT uses a voltage waveform, whilst the EMCRT uses a current waveform, as follows:

- a. **The ESCRT Waveform.** Fig 15 shows the voltage waveform required to produce a type A time base. The frequency of the waveform is equal to the pulse recurrence frequency of the radar. A circular time base, rotating with angular velocity  $\omega$ , can be obtained by applying voltages proportional to  $\sin \omega t$  and  $\cos \omega t$  to the X and Y plates respectively.

14-12 Fig 15 A-scope Waveform



- b. **The EMCRT Waveform.** As the coils of the EMCRT are subject to inductance, current modulations must be used to produce the sawtooth waveform. A PPI display can be produced by physically rotating the coils.

55. **Calibration Markers.** For accurate measurement of range, the time base of a CRT must be set up against some accurate standard. This is done by a calibrator unit which is synchronized with the transmitter pulse. The calibration markers appear as pips or bright rings.

56. **Types of Display.** The CRT in a radar receiver can be used to present one, two or three-dimensional information concerning the target.

- a. **One-dimensional Displays.** If only one-dimensional information is presented, the trace is deflection-modulated (ie the spot is deflected from its normal path to indicate the presence of an echo signal from a target). Examples of this type of display are the A-scope, I-scope, and J-scope.
- b. **Two-dimensional Displays.** For two-dimensional displays, the trace is intensity-modulated (ie the brilliance of the trace is varied by an echo signal from the target). Examples of two-dimensional displays are:
- (1) The PPI, sector-PPI, and B-scope, each of which shows the situation in a bearing and range plan format.
  - (2) The C-scope, which shows elevation and azimuth error, and is therefore of special use in fighters, where the display corresponds to the pilot's view through the windscreen. The C-scope information can also be readily transferred to the pilot's Head-up Display.



c. **Three-dimensional Displays.** Three-dimensional information is displayed in a variety of ways, often by combining complementary two-dimensional displays. For example, a C-scope, showing elevation and azimuth error, can be superimposed on to a B-scope showing range and azimuth. The I-scope shows target range, displacement from the axis, and angle-off error.

In all types of display, the spot moves in some pre-determined manner on the screen, this movement being termed the time base. Some of the commonly-used radar displays are illustrated in Volume 11, Chapter 1.

57. **Strobes.** In some radar equipments, it is necessary to expand a section of the main time base on either side of the echo blip in order to measure the range more accurately. To do this, a strobe pulse is generated at some definite time after the start of the main time base, the delay being controlled by the operator. The strobe pulse is made to trigger a strobe time-base circuit, with its own calibration on a larger scale.

## CHAPTER 13 - POWER SUPPLIES

Contents	Page
Introduction .....	2
BATTERIES.....	2
The Simple Cell .....	2
The Dry Cell.....	2
Mercury Cells.....	3
Solar Cells .....	4
Non-rechargeable Batteries .....	4
Rechargeable Batteries.....	4
Lead Acid Batteries .....	4
Nickel Based Batteries .....	7
Lithium Based Batteries .....	8
Lithium-polymer Battery.....	9
DC POWER FROM AC SOURCES .....	10
Overview of Process.....	10
Transformation .....	10
Rectification .....	10
Filters or Smoothing Circuits .....	12
Stabilization .....	13

### Table of Figures

14-13 Fig 1 An Early Dry Cell .....	3
14-13 Fig 2 Mercury Cells (Flat and Cylindrical) .....	3
14-13 Fig 3 The Lead-acid Battery.....	5
14-13 Fig 4 Component Parts of a Power Supply.....	10
14-13 Fig 5 Rectification.....	10
14-13 Fig 6 Types of Rectification.....	11
14-13 Fig 7 Filter/Smoothing Action .....	12
14-13 Fig 8 Rectifier Output with Capacitor .....	12
14-13 Fig 9 Rectifier Output with Capacitor and Load .....	13
14-13 Fig 10 Zener Diode .....	13

### Tables

Table 1 The Advantages and Limitations of Lead Acid Batteries .....	6
Table 2 The Advantages and Limitations of AGM Batteries .....	6
Table 3 The Advantages and Limitations of NiCd Batteries .....	7
Table 4 The Advantages and Limitations of Li-ion Batteries.....	8
Table 5 Six Common Lithium-ion Batteries.....	9

## Introduction

1. In order to function, all electronic equipment needs to be energized by means of a power supply. In the great majority of cases, this power needs to be delivered at a steady or fixed voltage. In the early days of radio and electronics, power was derived from batteries, but with the advent of thermionic valves, which called for high currents and voltages, this source of power became inconvenient. Equipment power supplies were therefore developed which relied on the main domestic supply as the energy source. With the invention of the transistor, battery packs came back into favour, and remain so due to modern digital technology allowing for equipment portability and modest power requirements. For larger power requirements or fixed installations mains or AC generator-driven power packs are usually more economical. For remote locations, such as space satellites, the solar cell has been developed to convert solar energy into electrical power.

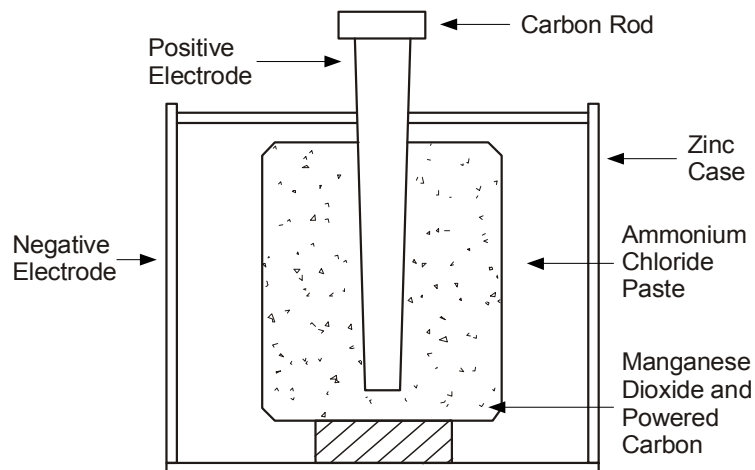
## BATTERIES

### The Simple Cell

2. When a metal electrode is immersed in an electrolyte, either:
  - a. Atoms of the electrodes of the more reactive metals (like zinc) are dissolved in the solution as positive ions, leaving the electrode negative with respect to the electrolyte, or
  - b. Positive ions from the electrolyte are attracted to the electrode of less reactive metals (like copper) making it positive with respect to the electrolyte.
3. If a zinc rod and a copper rod are placed in dilute acid and are joined by a wire, then the positive charge on the copper will tend to flow along the wire and neutralize the negative charge on the zinc. This will cause more zinc to dissolve in the acid and more hydrogen to be liberated at the copper so as to maintain the potential difference between the electrodes. Current will flow until all the zinc is dissolved away or until the acid is exhausted.
4. One disadvantage of the simple cell is that hydrogen bubbles being liberated at the copper electrode tend to insulate it from the solution and stop the chemical action. This effect is called polarization. Another disadvantage is that of 'local action'. Impurities present in the zinc form small local cells causing the zinc to be dissolved away without supplying power to the main circuit.

### The Dry Cell

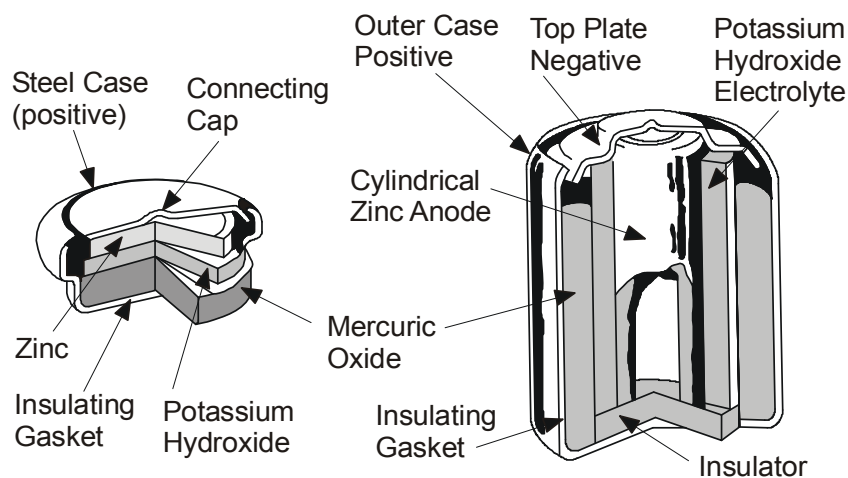
5. Fig 1 depicts an early dry cell, also known as a 'Leclanché cell'. The positive electrode is the carbon rod and the negative electrode is the zinc case. The electrolyte is ammonium chloride, in paste form, and is separated from the carbon rod by a depolarizing agent. This is a mixture of powdered manganese dioxide and carbon, which oxidizes the hydrogen molecules to water.

**14-13 Fig 1 An Early Dry Cell**

6. Similar materials were used in a later version of the Leclanché with the addition of mercuric chloride to reduce local action, and potassium dichromate to inhibit the corrosion of zinc. A disadvantage of this type of cell is that current quickly falls off, principally because hydrogen forms more quickly than it can be removed by the depolarizer. For this reason, they are best suited for intermittent work.

### Mercury Cells

7. The mercury cell, the full name of which is zinc-mercuric oxide, was developed during the Second World War for use in portable equipment where maximum energy with minimum volume was the prime aim. The modern day mercury cell, shown at Fig 2, is capable of providing a steady voltage over nearly all of the cell's useful discharge period. Over long periods of operational use, or long storage, a voltage regulation within one per cent of the initial voltage is still maintained.

**14-13 Fig 2 Mercury Cells (Flat and Cylindrical)**

8. In the mercury cell, the negative electrode is zinc, as in the Leclanché cell, but the positive electrode and its depolarizer consist of graphite and mercuric oxide. The electrolyte consists of potassium hydroxide.

### **Solar Cells**

9. The necessity for continuous electric power generation on space satellites led to the development of the solar cell. These devices, which are generally made from thin slices of highly pure single crystal silicon, produce electric power from radiant energy. Wavelengths in the range 400 to 1,100 nanometers are the most efficient, and about half the solar spectrum falls in this range.

10. Solar cells have no storage capacity, and for terrestrial applications they are used in conjunction with electrical storage batteries.

## **Non-rechargeable Batteries**

11. Considerable effort and research has been put into rechargeable batteries over recent years but non-rechargeable batteries continue to fill an important niche market in applications such as wristwatches, remote controls and electronic keys. Non-rechargeable batteries are useful when charging is impractical or impossible, such as in some military applications. They possess high specific energy, can be stored for protracted periods and be ready for immediate use. Most non-rechargeable batteries are relatively inexpensive to produce and environmentally friendly.

12. Carbon-zinc general purpose batteries are used for applications with low power drain such as remote controls and torches. One of the most common non-rechargeable batteries is the *alkaline-manganese*, or alkaline battery. Alkaline batteries deliver more energy at higher load currents than carbon-zinc and do not leak when depleted, as carbon-zinc does.

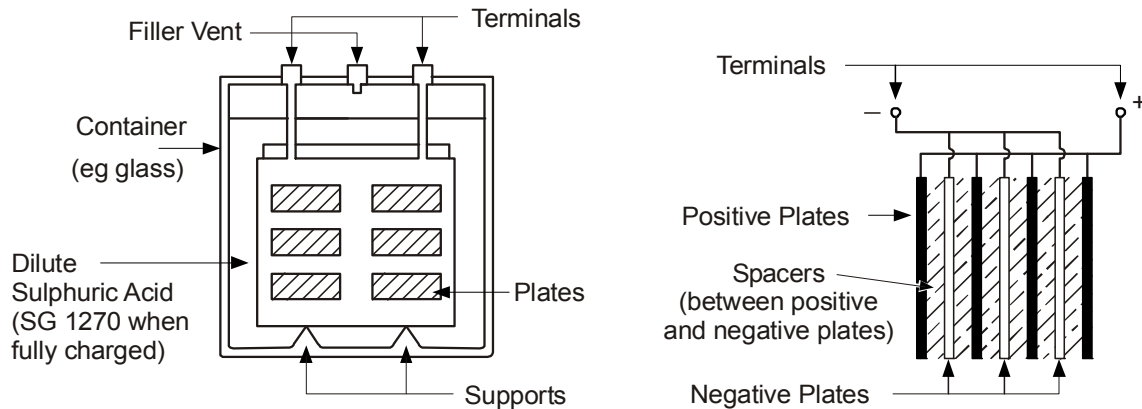
13. Non-rechargeable batteries have one of the highest energy densities. Although rechargeable batteries have improved, a regular household alkaline battery provides 50 percent more energy than a lithium-ion one. The most energy-dense non-rechargeable type is the lithium battery made for film cameras and military combat and holds over three times the energy of lithium-ion.

## **Rechargeable Batteries**

### **Lead Acid Batteries**

14. A lead-acid battery is a device for storing electricity. It produces power by means of chemical reaction. Unlike the batteries described so far, this reaction is reversible; so that when the battery has discharged power, it can be recharged by passing an electric current through it.

15. In its simplest form, shown in Fig 3, the lead-acid battery consists of two groups of interleaved plates, one group of lead and the other group of lead dioxide, immersed in a weak solution of sulphuric acid. Connecting the two groups of plates together electrically causes a chemical reaction to take place and current flows between the plates. This reaction continues until both plates change to lead sulphate, and the acid turns to water.

**14-13 Fig 3 The Lead-acid Battery**

16. The battery is recharged by passing an electric current (from a generator) through it in the other direction. This reverses all the electrical forces, and all the chemical reactions reverse themselves as a result. The lead sulphate changes back to lead oxide in the positive plates, and lead in the negative plates. The water changes back to sulphuric acid.

17. During the mid 1970s, researchers developed a maintenance-free lead acid battery that could operate in any position. The liquid electrolyte was transformed into moistened separators and the enclosure was sealed. Safety valves allow the venting of gas generated during charge and discharge.

18. Driven by different applications, two battery designations emerged. They are the small sealed lead acid (SLA), and the large valve regulated lead acid (VRLA). Technically, both batteries are the same. The term SLA is a misnomer as no lead acid battery can be totally sealed. As a result, the batteries are fitted with a valve to control the venting of gases during stressful charge and rapid discharge. Unlike the conventional lead acid battery where the plates are submerged in a liquid, the electrolyte is impregnated into a moistened separator. The design resembles nickel and lithium based systems and enables the battery to be operated in any physical orientation without leakage.

19. Both the SLA and VRLA are designed with a low over-voltage potential to prohibit the battery from reaching its gas-generating potential during charge. Excess charging would cause gassing and water depletion and consequently, these batteries should never be charged to their full potential.

20. Leaving the lead acid battery on float charge for a prolonged time does not cause damage. The battery's charge retention is the best among rechargeable batteries, for example, a Nickel-cadmium battery (NiCd) self-discharges approximately 40 percent of its stored energy in three months, the SLA self-discharges the same amount in one year. The SLA is relatively inexpensive but the operational costs can be higher than the NiCd if full cycles are required on a repetitive basis.

21. The SLA is not suitable for fast charging with typical charge times of 8 to 16 hours. The SLA must always be stored in a charged state as leaving the battery in a discharged condition causes sulphation, a condition that makes the battery difficult, if not impossible, to recharge.

22. Unlike the NiCd, the SLA is not suitable for deep cycling. A full discharge causes extra strain and each cycle robs the battery of a small amount of capacity. This loss is small while the battery is in good operating condition, but the fading increases once the performance drops to half the nominal capacity. This wear-down characteristic also applies to other battery chemistries in varying degrees.

23. Depending on the depth of discharge and operating temperature, the SLA provides 200 to 300 discharge/ charge cycles. The primary reason for its relatively short cycle life is grid corrosion of the positive electrode, depletion of the active material and expansion of the positive plates. These changes are most prevalent at elevated operating temperatures and high-current discharges.

24. The optimum operating temperature for the SLA and VRLA battery is 25 °C. As a rule of thumb, every 8 °C rise in temperature will cut the battery life in half. VRLA that would last for 10 years at 25 °C will only be good for 5 years if operated at 33 °C. The same battery would endure a little more than one year at a temperature of 42°C.

25. Among modern rechargeable batteries, the lead acid battery family has the lowest energy density, making it unsuitable for handheld devices that demand compact size. They work well at cold temperatures and are superior to lithium-ion when operating in subzero conditions.

**Table 1 The Advantages and Limitations of Lead Acid Batteries**

Advantages	Limitations
Inexpensive and simple to manufacture	Should not be stored in a discharged condition
Mature, reliable and well-understood technology	Low energy density
Durable and provides dependable service	Limited number of full discharge cycles
Low self-discharge	Environmentally unfriendly
Low maintenance requirements	Transportation restrictions on flooded lead acid
Capable of high discharge rates	Thermal runaway possible with improper charging
Good low and high temperature performance	

26. **Absorbent Glass Mat (AGM) Batteries.** AGM technology became popular in the early 1980s as a sealed lead acid battery for military aircraft, vehicles and uninterrupted power supplies to reduce weight and improve reliability. The acid is absorbed by a very fine fibreglass mat making the battery spill-proof. The plates can be made flat to resemble a standard flooded lead acid pack in a rectangular case or wound into a cylindrical cell. AGM has very low internal resistance, is capable of delivering high currents on demand and offers a relatively long service life, even when deep-cycled. AGM is maintenance free, provides good electrical reliability and is lighter than the flooded lead acid type. It stands up well to low temperatures and has a low self-discharge. The leading advantages are a charge that is up to five times faster than the flooded version, and the ability to deep cycle. As with all gelled and sealed units, AGM batteries are sensitive to overcharging.

**Table 2 The Advantages and Limitations of AGM Batteries**

Advantages	Limitations
Spill-proof	Higher manufacturing cost
High specific power, low internal resistance, responsive to load	Sensitive to overcharging
Up to 5 times faster charge than with flooded technology	Capacity has gradual decline
Better cycle life than with flooded systems	Low specific energy
Vibration resistance	Must be stored in charged condition
Stands up well to cold temperature	Not environmentally friendly

## Nickel Based Batteries

27. **Nickel-cadmium Cells.** Rechargeable nickel-cadmium cells are useful in electronic equipment since they can be sealed, thus avoiding the effect of corrosive fumes which are given off by lead-acid accumulators. The sealed type of cell has a long life and can be completely discharged without ill effects.

**Table 3 The Advantages and Limitations of NiCd Batteries**

Advantages	Limitations
Fast and simple charging after prolonged storage	Relatively low specific energy compared with newer systems
High number of charge/discharge cycles (over 1000 with proper care and maintenance)	Memory effect; needs periodic full discharges
Good load performance; rugged and forgiving if abused	Environmentally unfriendly; cadmium is toxic
Long shelf life; can be stored in a discharged state	High self-discharge; needs recharging after storage
Simple storage and transportation	
Good low-temperature performance	
Economical	
Wide range of sizes and performance options	

28. **Nickel-metal-hydride (NiMH).** NiMH provides 40 percent higher specific energy than a standard NiCd and does not contain toxic metals. NiMH also has a high self-discharge rate and loses about 20 percent of its capacity within the first 24 hours, and 10 percent per month thereafter. The NiMH battery in a torch in storage will be 'flat' after only a few weeks.

**9-12 Table 1 The Advantages and Limitations of NiMH Batteries**

Advantages	Limitations
30 % to 40 % higher capacity than a standard NiCd	Limited service life; deep discharge reduces service life
Less prone to memory effect than NiCd	Requires complex charge algorithm
Simple storage and transportation	Does not absorb overcharge well; trickle charge must be kept low
Environmentally friendly	Generates heat during fast-charge and high-load discharge
Nickel content makes recycling profitable	High self-discharge
	Performance degrades if stored at elevated temperatures; should be stored in a cool place at about 40 % charge



## Lithium Based Batteries

29. **Lithium (Li) Batteries.** There are two types of lithium battery. Lithium-ion batteries are rechargeable and used in equipment such as laptops, tablets and smart phones. Lithium-metal batteries are non-rechargeable and are found in items such as watches. Although lithium batteries are safe, their high energy levels mean they can pose a fire risk if damaged or charged incorrectly. Rechargeable batteries with lithium metal on the anode could provide high energy densities; however, cycling produces unwanted growth particles (dendrites) on the anode which can penetrate the separator and cause an electrical short. When this occurs, the cell temperature rises quickly and approaches the melting point of lithium causing thermal runaway, also known as “venting with flame”. As a result Li batteries must be treated with care and stowed appropriately during flight. Lithium batteries are known to have been the cause of a number of fires onboard aircraft and during ground handling. Additionally, poor quality or counterfeit batteries, which are known to be in wide circulation, pose an additional safety risk.

30. **Lithium Ion (Li-ion) Batteries.** The inherent instability of lithium metal, especially during charging has resulted in the production of a non-metallic solution using *lithium ions*. Although lower in specific energy than lithium-metal Li-ion is safe provided manufacturers and battery packers follow the correct safety measures in limiting voltages and currents to secure levels. The specific energy of a Li-ion battery is twice that of NiCd and it is also low-maintenance. The battery has no memory and does not need exercising (deliberate full discharge). Self-discharge is less than half that of nickel-based systems. The disadvantages of the LI-ion battery are the need for protection circuits to prevent abuse and high cost.

31. Similar to lead and nickel based architectures, lithium-ion uses a cathode (positive electrode) made from a metal oxide, an anode (negative electrode) made of porous carbon and electrolyte as the conductor. During discharge, the ions flow from the anode to the cathode through the electrolyte and separator. Charging reverses the direction and the ions flow from the cathode to the anode. Li-ion batteries come in many varieties which vary in performance where the choice of cathode material determines the unique characteristics of the battery. Most manufacturers use graphite as the anode to attain a flatter discharge curve. Graphite stores lithium-ions effectively when the battery is charged and has long-term cycle stability.

**Table 4 The Advantages and Limitations of Li-ion Batteries**

Advantages	Limitations
High energy density	Requires protection circuit to limit voltage and current
Relatively low self-discharge; less than half that of NiCd and NiMH	Subject to aging, even if not in use
Low maintenance	Transportation regulations when shipping in larger quantities
No periodic discharge is needed	
No memory	

**Table 5 Six Common Lithium-ion Batteries**

<b>Li-ion Battery</b>	<b>Characteristics</b>
Lithium Cobalt Oxide( $\text{LiCoO}_2$ )	High specific energy / Relatively short life span and limited load capabilities (specific power).
Lithium Manganese Oxide ( $\text{LiMn}_2\text{O}_4$ )	Low internal cell resistance - fast charging and high-current discharging / Capacity that is roughly one-third lower compared to Li-cobalt.
Lithium Iron Phosphate( $\text{LiFePO}_4$ )	Enhanced safety, good thermal stability, tolerant to abuse, high current rating and long cycle life / cold temperature reduces performance, and elevated storage temperature shortens the service life.
Lithium Nickel Manganese Cobalt Oxide ( $\text{LiNiMnCoO}_2$ )	Can also be tailored to high specific energy or high specific power, but not both.
Lithium Nickel Cobalt Aluminium Oxide ( $\text{LiNiCoAlO}_2$ )	High specific energy and power densities / long life span.
Lithium Titanate ( $\text{Li}_4\text{Ti}_5\text{O}_{12}$ )	Li-titanate replaces the graphite in the anode / can be fast-charged and delivers a high discharge current / High cycle rate / Safe / Excellent low-temperature discharge characteristics.

**Lithium-polymer Battery**

32. Lithium-polymer differs from other battery systems in the type of electrolyte used. The original polymer design used a solid (dry) polymer electrolyte that resembles a plastic-like film. This insulator allows the exchange of ions and replaces the traditional porous separator that is soaked with electrolyte. A solid polymer has a poor conductivity at room temperature and the battery must be heated to between 50 and 60 °C (122 to 140 °F) to enable the current to flow.

33. To make the Li-polymer battery conductive at room temperature, gelled electrolyte is added. All Li-ion polymer cells today incorporate a micro porous separator with moisture. The correct term is "Lithium-ion polymer" (Li-ion polymer or Li-polymer for short). Li-polymer can be built on many systems, such as Li-cobalt, Li-phosphate and Li-manganese. For this reason, Li-polymer is not considered a unique battery chemistry.

34. Li-polymer cells may also be shaped into a flexible foil-type case (polymer laminate or pouch cell) that resembles a food package. While a standard Li-ion needs a rigid case to press the electrodes together, Li-polymer uses laminated sheets that do not need compression. Although less durable than conventional packaging, a foil-type enclosure reduces the weight by more than 20 percent over the classic hard shell and thin film technology liberates the format design and the battery can be made into any shape.

35. Charge and discharge characteristics of Li-polymer are identical to other Li-ion systems and do not require a special charger. Safety issues are also similar in that protection circuits are needed. Gas build-up during charge can cause some Li-polymer in a foil package to swell. Li-polymer is not limited to a foil package shape and can also be made into a cylindrical design.

36. **Li-cobalt.** Most Li-polymer packs for the consumer market are based on Li-cobalt. With gelled electrolyte added the lithium polymer is essentially the same as the lithium-ion battery. Both

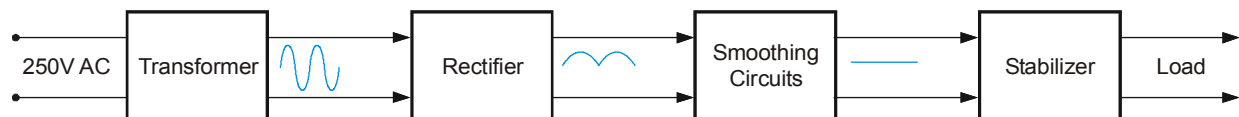
use identical cathode and anode material and contain a similar amount of electrolyte. Although the characteristics and performance of the two systems are alike, the Li-polymer is unique in that a micro porous electrolyte replaces the traditional porous separator. The gelled electrolyte becomes the catalyst that enhances the electrical conductivity. Li-polymer offers slightly higher specific energy and can be made thinner than conventional Li-ion.

## DC POWER FROM AC SOURCES

### Overview of Process

37. When comparatively large amounts of power are needed, the sources of supply are nearly always alternating currents. However, before the power can be used inside a piece of electronic equipment it needs to be converted to a DC of the correct magnitude. The component parts of this process are shown in Fig 4.

**14-13 Fig 4 Component Parts of a Power Supply**



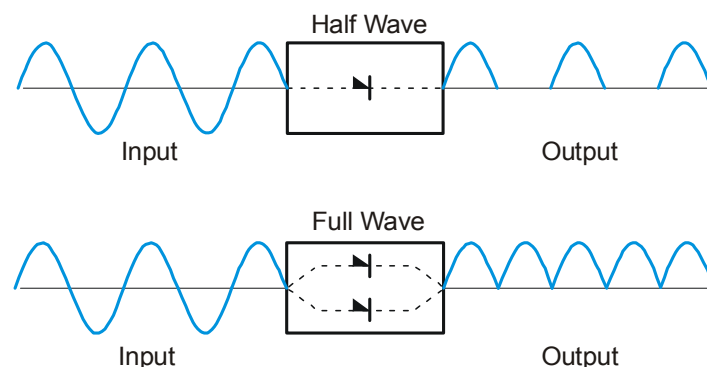
### Transformation

38. In most cases, the amplitude of the AC voltage delivered to a piece of electronic equipment is of the incorrect value and needs to be transformed, up or down, to suit the equipment needs. This transformation is accomplished by means of a transformer. The subject of transformers is covered in some detail in Volume 14, Chapter 11.

### Rectification

39. Once the AC voltage has been transformed to a level that is suitable for use in the equipment, the first stage of converting AC into a DC takes place. This stage is known as rectification, and involves the production of a pulsating unidirectional output from an alternating input. The process can be either half-wave rectification or full-wave rectification (see Fig 5).

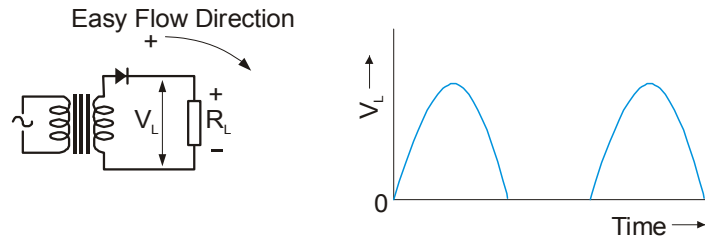
**14-13 Fig 5 Rectification**



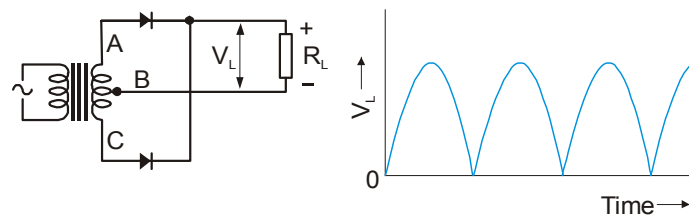
40. **Half-wave Rectification.** The diode (thermionic or semi-conductor) is the fundamental component in all rectification circuits. As the diode conducts only when the anode is positive with respect to the cathode, pulses of current flow through the load on alternate half-cycles of the voltage waveform. The output of a half-wave rectifier is shown in Fig 6a.

14-13 Fig 6 Types of Rectification

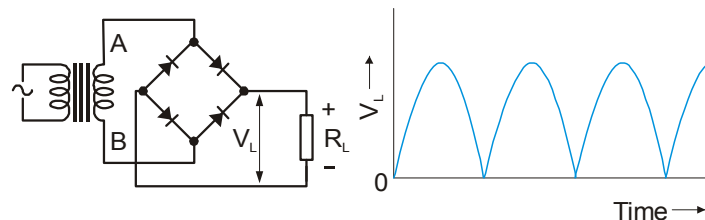
**a Half-wave Rectifier**



**b Full-wave Rectifier**



**c Bridge Rectifier**



41. **Full-wave Rectification.** By using two diodes, the full applied waveform can be used to produce an output, as shown in Fig 6b. The anodes of the two diodes are connected to opposite ends of a centre-tapped transformer winding. The voltages at the diode anodes are thus in anti-phase, so that the diodes conduct on alternate half-cycles of the applied waveform. The current through the load is always in the same direction, regardless of which diode is conducting, so producing a unidirectional flow of pulses through the load on every half-cycle. Note that a double diode could be used instead of two separate diodes.

421. **Bridge Rectification.** An alternative form of a full-wave rectifier is the bridge rectifier. This does away with the need for a centre-tapped transformer, but uses four diodes (thermionic or semi-conductor) arranged like a Wheatstone Bridge, with the supply connected across one diagonal of the bridge and the load across the other. The diodes act in pairs and, as before, the current through the load is always in the same direction (Fig 6c).

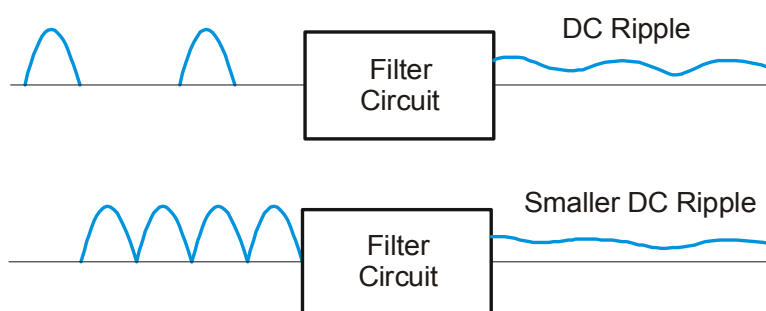
43. **P-N Junction Diode as a Rectifier.** Since a junction diode passes current only when the voltage across it is of one polarity and allows almost none to flow when the polarity is reversed, it may be used as a rectifier.

44. **Practical Junction Diode Rectifiers.** Junction diodes are manufactured in a wide range of sizes and cases. In general, the more power a diode has to handle, the bigger it will be. If the diode is passing a large current, heat is developed because of the resistance of the diode. This heat must be removed if the diode is not to be damaged, and the easiest way to do this is to provide large surface areas exposed to the air. Very often special 'heat sinks' are provided to give quick removal of heat from the diode.

### Filters or Smoothing Circuits

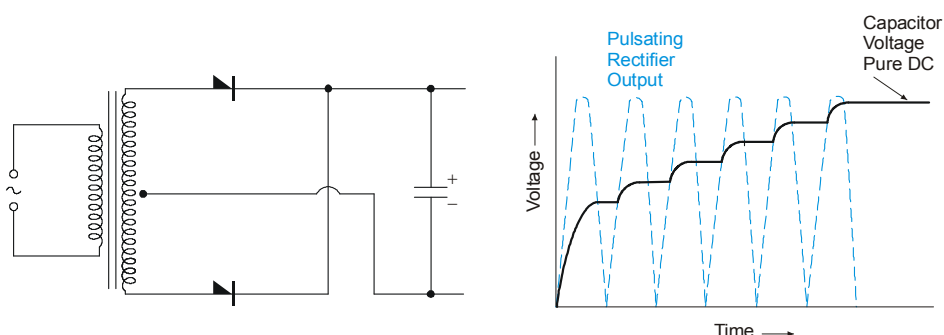
45. The fluctuating output from the rectification stage cannot be regarded as a unidirectional supply suitable for equipment use; the AC component needs to be minimized. This is achieved by a filter, or smoothing circuit and the action is illustrated in Fig 7.

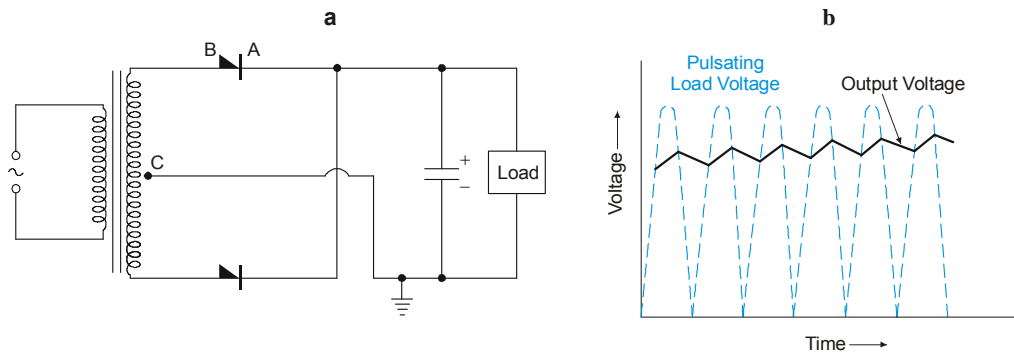
14-13 Fig 7 Filter/Smoothing Action



46. **Reservoir Capacitor.** If a large value capacitor is placed across the output terminals of a rectifier, and the load is disconnected, a steady DC voltage is developed across the capacitor. A rectifier has resistance, due mainly to the internal resistance of the rectifying device, therefore, when a capacitor is placed across the output of a rectifier, it charges up every time the rectifier conducts. If it does not reach the full output voltage at the end of the first half-cycle, it will do so during the next few half-cycles. It cannot discharge between the peaks of the pulsating DC because the rectifier will not pass current in the reverse direction. Therefore, after a few half-cycles, a steady DC voltage appears across the capacitor, as shown in Fig 8, and the AC has been filtered out. If a load is connected across the capacitor (Fig 9a), the capacitor begins to discharge and its voltage will fall. However, in a full-wave rectifier a new voltage peak occurs 100 times per second and this voltage peak recharges the capacitor. The resulting output waveform is shown in Fig 9b. The DC is not quite steady, but the amplitude of the pulses is much smaller than if there were no capacitor at all. The voltage waveform is now DC plus a small component of AC called the 'ripple'. The voltage is not smooth enough for the final HT output and more components are needed in the smoothing circuit.

14-13 Fig 8 Rectifier Output with Capacitor



**14-13 Fig 9 Rectifier Output with Capacitor and Load**

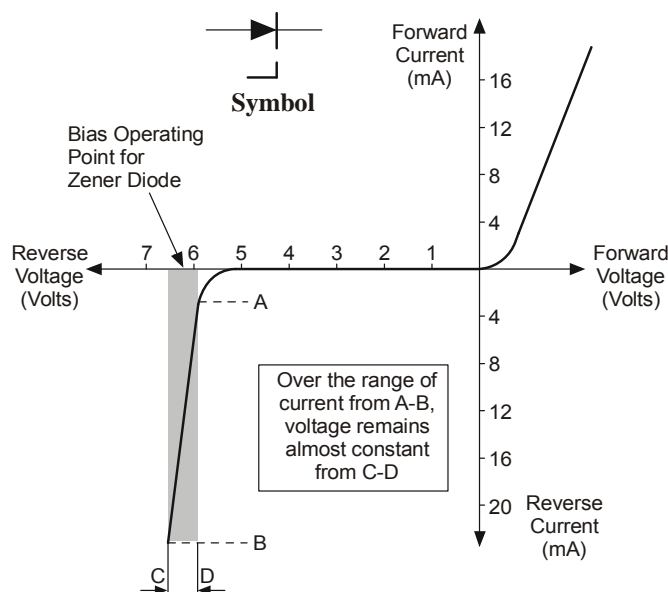
47. **Smoothing Circuit Resistor.** The ripple can be reduced by adding a resistor and capacitor, which will act as an AC voltage divider, to the smoothing circuit. It is common for two resistors to be connected across the capacitor. This would normally give constant voltage to within 1 to 2% of the required HT supply. If this is not satisfactory, another RC section could be added.

### Stabilization

48. If it is required to ensure that the DC supply from the smoothing circuits remains constant, regardless of the applied load, compensation is needed. This compensation is generally known as stabilization.

49. Although there are a number of such circuits, they are usually based around the cold-cathode, gas-filled diode for HT circuits and the zener diode for low voltage circuits.

50. **Zener Diode.** Most junction diodes will be permanently damaged by the high reverse current flowing if the breakdown voltage is exceeded. The zener diode is specially designed to make use of this part of the diode characteristic, and is not damaged by the high reverse current. If the zener diode is reverse biased to a value just beyond the breakdown voltage, as shown in Fig 10, the voltage across the diode remains practically constant for large variations of current through it. Therefore, the zener diode may be used to provide a constant reference voltage point in a circuit, or as a means of stabilizing the voltage over a wide range of circuit currents. A whole range of zener diodes with different values of breakdown (stabilizing) voltage is available.

**14-13 Fig 10 Zener Diode**

## CHAPTER 14 - AMPLIFIERS

Contents	Page
Introduction .....	2
The Transistor as an Amplifier .....	2
Gain .....	4
Transistor Amplifier Action .....	4
Transistor Circuit Arrangements.....	5
Amplifiers in Cascade.....	6
Resistance Capacitance (RC) Coupling.....	6
Transformer Coupling.....	7
Decibels.....	8
Video Amplifiers.....	11
Radio Frequency (RF) Amplifiers.....	12
Direct-coupled Amplifiers .....	12
AMPLIFIERS WITH FEEDBACK .....	13
Introduction.....	13
Principle of Negative Feedback.....	13
The Effect of Feedback on Frequency Response.....	14
Frequency Selective Feedback .....	14
Input and Output Impedances with Feedback.....	14
POWER AMPLIFIERS.....	15
Introduction.....	15
Classes of Operation.....	15

### Table of Figures

14-14 Fig 1 Transistor Principles.....	3
14-14 Fig 2 Transistor Amplifier Circuit.....	3
14-14 Fig 3 Transistor Amplifier Action .....	5
14-14 Fig 4 Three Circuit Arrangements.....	5
14-14 Fig 5 Cascade-connected Amplifiers .....	6
14-14 Fig 6 Coupling Network.....	7
14-14 Fig 7 Transformer Inter-stage Coupling .....	7
14-14 Fig 8 Comparison Between Transformer and RC Couplings .....	8
14-14 Fig 9 Gain in a Multi-stage Amplifier .....	8
14-14 Fig 10 Gain Frequency Curve.....	9
14-14 Fig 11 Conversion to Decibels from Voltage and Power Ratios .....	11
14-14 Fig 12 Ideal Video Wave-forms.....	12
14-14 Fig 13 Block Diagram Representation of a Feedback Amplifier .....	13
14-14 Fig 14 Frequency Response Curves of an Amplifier with and without Negative Feedback .....	14

### Tables

Table 1 Transistor Circuit Comparisons.....	6
Table 2 Changes to Input and Output Impedances due to Feedback .....	14

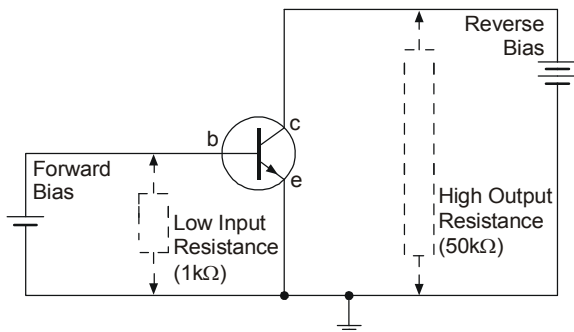
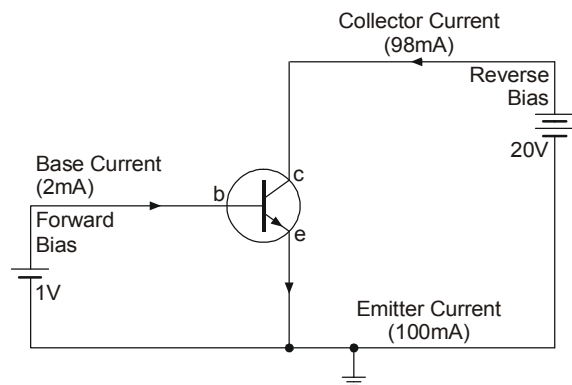
## Introduction

1. Amplification is needed in a system in order to boost weak signals to a level where they can be of some practical use. In the case of a radio receiver, signals at the aerial, of the order of a few microwatts, require amplification before they can be used to drive a speaker system demanding several watts.
2. Amplifiers are classified according to the function they perform. There are four main types of amplifier, each designed to amplify signals in a specific part of the frequency spectrum:
  - a. **Audio Amplifiers.** These amplify a band of frequencies in the audio range, ie between about 20 and 20,000 Hz. Audio amplifiers are used in radio and television receivers and transmitters, in intercom equipment, CDs, cassette tape decks, and in many other electronic devices.
  - b. **Video Amplifiers.** These are similar to audio amplifiers except that they cover a much wider frequency band, ranging from about 20 Hz to 6 MHz. Whereas audio amplifiers deal with electrical waveforms corresponding to sounds, video amplifiers handle electrical waveforms corresponding to pictures. They are used in television and in radar apparatus.
  - c. **RF Amplifiers.** These amplify only a narrow band of frequencies, so narrow that they may be considered to amplify at one frequency only. This frequency may be anywhere in the RF band, ie from 20 kHz to 100 GHz. They are used in radio and television transmitters and receivers, in radar and in weapon guidance systems.
  - d. **DC Amplifiers.** These direct-coupled amplifiers are used to amplify inputs at frequencies of a few hertz and slow non-regular changes in voltage. They are not used to amplify steady non-varying DC, as this would be pointless. They are used in many types of electronic equipment, particularly in analogue computers.
3. These main categories for amplifiers may be broken down further into the sub-categories of wideband, narrow band, tuned, stagger tuned, voltage and power. Thus, for example, a particular amplifier may be referred to as an AF stagger tuned voltage amplifier.

## The Transistor as an Amplifier

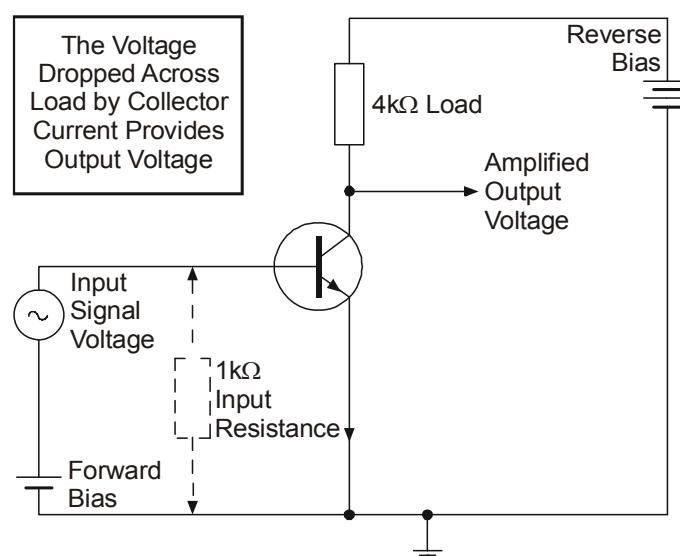
4. In a properly biased transistor the resistance of the forward biased emitter-base (input) circuit is low whilst that of the reverse biased collector (output) circuit is high. Typical values for the arrangement shown in Fig 1 are one k $\Omega$  and 50 k $\Omega$  respectively.
5. The forward and reverse-bias voltages are such that certain steady values of current flow in the emitter, base and collector circuits. Ninety-eight percent of the electrons from the emitter reach the collector; therefore if the emitter current is 100 mA, the collector current will be about 98 mA, and the base current about 2 mA as shown in Fig1b. Furthermore, the ratios of these currents remain constant so that if one current changes the other two also change. For example, using the ratios given above, if the emitter current changes by 1 mA, the collector current will change by 0.98 mA and the base current by 0.02 mA. Similarly, if the base current is made to change by 0.02 mA, the collector current will change by 0.98 mA and the emitter current by 1 mA. In an amplifier, it is these changes in currents and voltages which generate interest.



**14-14 Fig 1 Transistor Principles****a. Input and Output Resistances****b. Division of Current**

6. If a very low resistance alternating voltage source in the emitter-base circuit is connected in series with the forward bias battery, the voltage applied to the input circuit can be varied. A  $4\text{k}\Omega$  load resistor in the collector circuit is also connected in series with the reverse bias battery as shown in Fig 2. The alternating voltage source represents an input signal voltage, eg from a pick-up on a record player. The output voltage is taken between the collector and earth.

7. If the input signal voltage in the circuit of Fig 2 is changed from zero to  $0.02\text{ V}$ , the base current flowing in the  $1\text{ k}\Omega$  input resistance will change by  $0.02\text{ mA}$  (ie  $0.02\text{V}/1\text{ k}\Omega$ ). This causes the collector current to change by  $0.98\text{ mA}$ .

**14-14 Fig 2 Transistor Amplifier Circuit**

## Gain

8. Technically, gain is defined as the increase in power level in the load, ie the ratio of the actual power delivered to that which would be delivered if the source were correctly matched, without loss, to the load, in the absence of an amplifier. Put more simply, gain in an amplifier is the ratio of output quantity to input quantity. Therefore, using the previous paragraph figures:

$$\begin{aligned}\text{Current gain} &= \frac{\text{Output (collector) current change}}{\text{Input (base) current change}} \\ &= \frac{0.98 \text{ mA}}{0.02 \text{ mA}} = 49\end{aligned}$$

9. The change in output current flowing in the 4 kΩ load resistor produces a change of voltage across the load. This change of voltage represents the output signal voltage, given by  $IR = 0.98 \text{ mA} \times 4 \text{ k}\Omega = 3.92 \text{ V}$ . Therefore:

$$\text{Voltage gain} = \frac{\text{Output voltage change}}{\text{Input voltage change}} = \frac{3.92 \text{ V}}{0.02 \text{ V}} = 196$$

10. There is also a power gain given by:

$$\text{Current gain} \times \text{voltage gain} = 49 \times 196 = 9,604$$

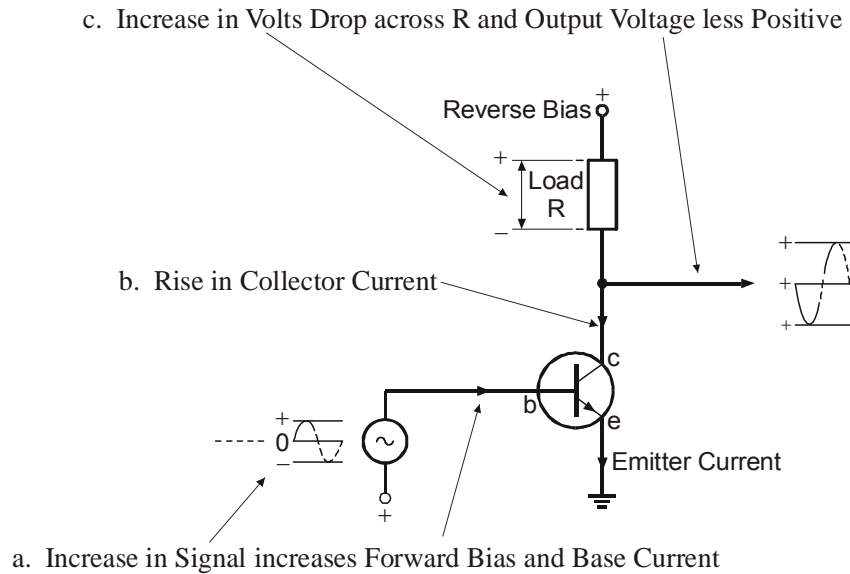
11. All the output quantities are greater than the input quantities and amplification has been achieved. The large voltage and power gains arise from the fact that the transistor transfers current from a low resistance input circuit to a high resistance output circuit.

## Transistor Amplifier Action

12. The action of the circuit of Fig 2 is summarized in Fig 3:

- a. The increase in input signal voltage increases the forward bias at the emitter-base junction (battery and signal voltages add).
- b. The increase in forward bias increases the base current and this causes the collector current to rise also - much more than the base current.
- c. The rise in collector current increases the voltage drop across the load resistor and this causes the output (collector-to-earth) voltage to become smaller (less positive).
- d. Note the phase relationship between input and output signals. An increase in input signal voltage causes a fall in output voltage. Thus, in this circuit, input and output voltages are 180° out of phase.

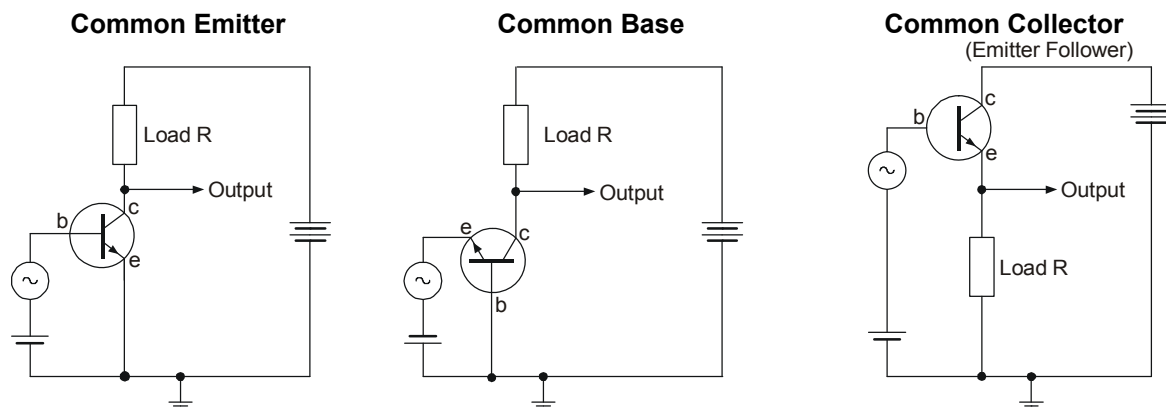
Similar results would have been obtained if the above had been applied to a p-n-p transistor amplifier.

**14-14 Fig 3 Transistor Amplifier Action**

13. The transistor amplifier action has also been covered in some detail, as it is imperative that the principles are understood.

### Transistor Circuit Arrangements

14. In the transistor amplifier circuits so far discussed, the emitter is common to both input and output circuits. This arrangement is called a common emitter or grounded emitter amplifier. Although the common emitter amplifier is the arrangement used most often, there are other ways of connecting the input and output circuits to the transistor. In practice, three circuit arrangements are in use as shown in Fig 4.

**14-14 Fig 4 Three Circuit Arrangements**

15. When dealing with the common emitter circuit several important factors are noted:

- The input resistance is low and the output resistance high.
- The current, voltage and power gains are all high.
- The input and output signals are  $180^\circ$  out of phase.

16. If this procedure is repeated for the other two circuits, several points of difference will be noted. Table 1 compares the important factors for each of the three circuits. Each arrangement is used only for those jobs to which it is suited.

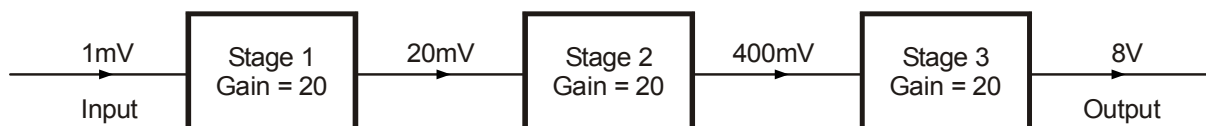
**Table 1 Transistor Circuit Comparisons**

	<b>Common Emitter</b>	<b>Common Base</b>	<b>Common Collector</b>
Input Resistance	Low (1 k $\Omega$ )	Very low (100 $\Omega$ )	High (100 k $\Omega$ )
Output Resistance	High (50 k $\Omega$ )	Very high (250 k $\Omega$ )	Low (200 $\Omega$ )
Current Gain	High (50)	Low (less than 1)	High (50)
Voltage Gain	High (50-1,000)	High (50-1,000)	Low (less than 1)
Power Gain	High (100-10,000)	Moderate (10-1,000)	Low (10-100)
Phase Relationship	180° Out of phase	In phase	In phase

### Amplifiers in Cascade

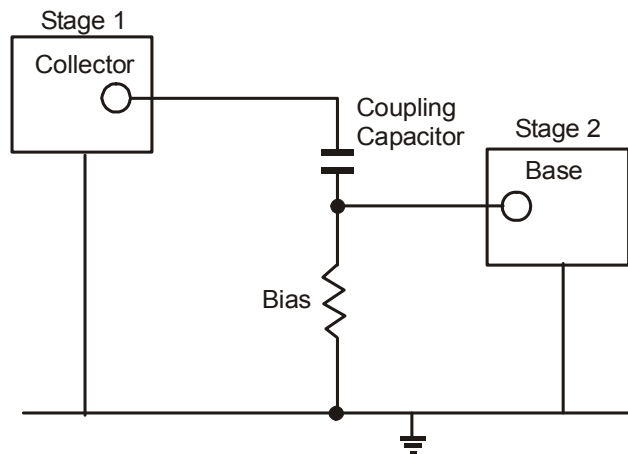
17. So far the amplifying device has been considered as a single amplifier stage. In some cases the amplification afforded by one stage is sufficient, but more often several stages are needed, one following after the other, to obtain the required amplification (Fig 5). Therefore, if in a three-stage amplifier each stage has a gain of 20 and the input to the first stage is 1 mV, the output from this stage is 20 mV. This is applied as the input to the second stage which gives an output of 400 mV. This becomes the input to the third stage which provides an output of 8 volts. The overall amplification of the three stages is  $20 \times 20 \times 20 = 8,000$ . The three amplifier stages are said to be connected in cascade.

**14-14 Fig 5 Cascade-connected Amplifiers**



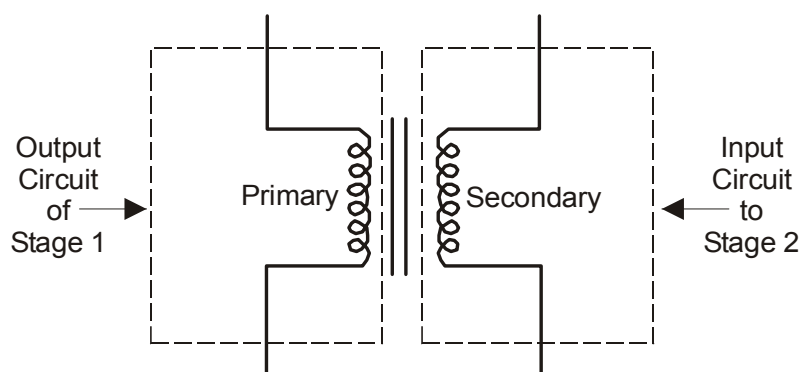
### Resistance Capacitance (RC) Coupling

18. In order to couple two amplifier stages together some form of connection between the output of the first stage and the input to the next stage is required. A direct connection between the collector of stage one and the base of stage two, would not do, since the collector DC voltage would be applied to the base and would upset the biasing arrangements. In order to block this DC and still allow the AF variations of collector voltage to be applied to the input of the next stage, a coupling capacitor is needed (Fig 6). This capacitor and the bias, form a potential divider across the output of stage one. All the DC voltage is dropped across the capacitor and most of the AF voltage is developed across the resistor and applied to the input of the next stage. This is a common method of coupling the stages of an AF amplifier.

**14-14 Fig 6 Coupling Network**

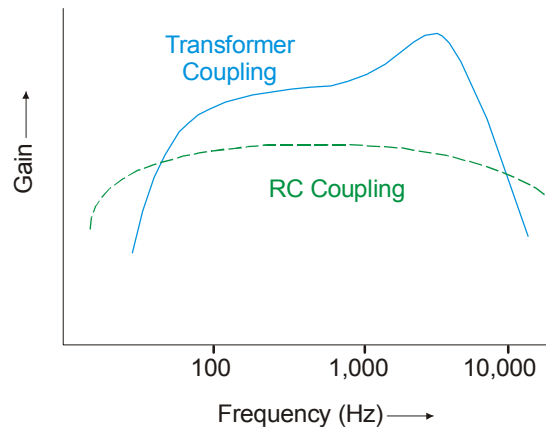
### Transformer Coupling

19. Another method of coupling the stages of an AF amplifier is by using transformers. The primary winding of the AF transformer forms the output load of the first stage and the secondary winding forms the input circuit of the next stage (Fig 7). No coupling capacitor is required.

**14-14 Fig 7 Transformer Inter-stage Coupling**

20. With transformer coupling, a higher stage gain can be obtained. In a transistor amplifier, by choosing the correct turns ratio, the high output impedance of stage one can be matched to the low input impedance of stage two thus giving a considerable increase in gain over RC-coupled stages.

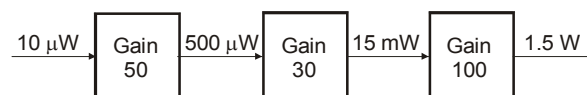
21. The main disadvantage of transformer coupling is the poor frequency response compared with an RC-coupled amplifier. The impedance of the windings varies with frequency; thus at the bottom end of the frequency range the impedance of the primary winding is low and so the voltage developed across it is low. At the high end of the frequency range the impedance is high and so these frequencies receive greater amplification than the middle frequencies. The response drops sharply at the top end of the frequency range due to the shunting effect of stray capacitances. The frequency response curves of transformer and RC-coupled amplifiers are compared in Fig 8. The uneven response of transformer coupling results in distortion.

**14-14 Fig 8 Comparison Between Transformer and RC Couplings**

22. For this reason, and because of the size and weight of iron-cored AF transformers, RC inter-stage coupling is preferred. Because they provide a simple way of matching a high impedance to a low impedance, transformers are often used between the output stage and a loudspeaker load.

### Decibels

23. The gain of an amplifier (see para 8) is the output divided by the input. We have so far expressed the gain as a number without units and as can be seen from Fig 9 in a multi-stage amplifier this number can become quite large.

**14-14 Fig 9 Gain in a Multi-stage Amplifier**

24. Radio amplifiers commonly have a power gain of several million and the overall gain figure becomes very cumbersome. Thus it is more convenient to express the power gain of an amplifier as the logarithm, to the base 10, of the output to input power ratio. The basic unit for this measurement is the bel. Therefore:

$$\text{Power gain} = \log_{10} \frac{\text{Power}_{\text{out}}}{\text{Power}_{\text{in}}} \text{ bels}$$

In practice the bel is a very large unit, so work is done in decibels (dB), a decibel being one-tenth of a bel. Therefore:

$$\text{Power gain in decibels} = 10 \log_{10} \frac{\text{Power}_{\text{out}}}{\text{Power}_{\text{in}}} \text{ dB}$$

25. The amplifier in Fig 9 would have a power gain, expressed in dB, of:

$$10 \log_{10} 50 + 10 \log_{10} 30 + 10 \log_{10} 100 \text{ dB} = (10 \times 1.699) + (10 \times 1.477) + (10 \times 2) = 51.76 \text{ dB}$$

Note that when the gain of each stage is expressed in decibels, the overall gain of the amplifier is found by adding the gain of individual stages.

26. In the case of an attenuator the output power is less than the input power, the power loss being indicated by a minus sign, e.g.

Input power = 160 watts

Output power = 5 watts

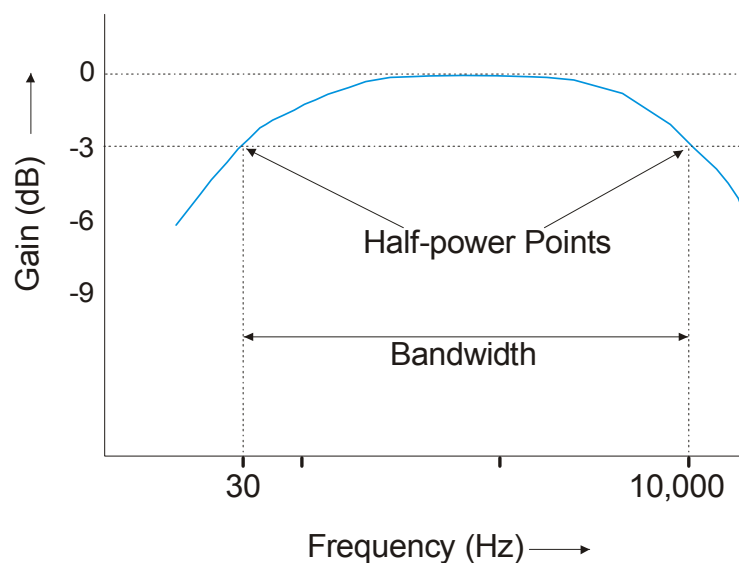
$$\begin{aligned}
 \text{Attenuation in dB} &= 10 \log_{10} \frac{\text{Power}_{\text{out}}}{\text{Power}_{\text{in}}} \\
 &= -10 \log_{10} \frac{\text{Power}_{\text{in}}}{\text{Power}_{\text{out}}} \\
 &= -10 \log_{10} \frac{160}{5} \\
 &= -15 \text{ dB}
 \end{aligned}$$

27. Fig 10 shows the frequency response curve for an amplifier; the gain is measured in dB. At the lower and upper ends of the curve where the response falls off, the gain can be expressed as minus so many decibels. A commonly used point is where the power has fallen to half the normal value. Therefore:

$$\begin{aligned}
 \text{Power loss} &= 10 \log \frac{1}{2} = -10 \log 2 \\
 &= -3\text{dB, or 3dB down}
 \end{aligned}$$

This point is known as the half-power point and the frequency range between the two half-power points is known as the bandwidth of the amplifier.

**14-14 Fig 10 Gain Frequency Curve**



28. With voltage amplifiers it is more convenient to express the gain of an amplifier in terms of a voltage ratio.

$$\text{Power} = \frac{V^2}{R}$$

$$\text{Power gain} = \frac{V_{\text{out}}^2}{R_{\text{out}}} \div \frac{V_{\text{in}}^2}{R_{\text{in}}}$$

$$= \frac{V_{\text{out}}^2}{V_{\text{in}}^2} \times \frac{R_{\text{in}}}{R_{\text{out}}}$$

If the amplifier has the same value of input and output resistances:

$$\text{Power gain} = \frac{V_{\text{out}}^2}{V_{\text{in}}^2}$$

Using decibels the power gain becomes:

$$\text{Power gain} = 10 \log \left( \frac{V_{\text{out}}}{V_{\text{in}}} \right)^2 \text{ dB}$$

$$= 20 \log \frac{V_{\text{out}}}{V_{\text{in}}} \text{ dB}$$

Similarly, since power also equals  $I^2 R$ , gain can be expressed in terms of a current ratio provided that the values for  $R$  in input and output are equal. Therefore:

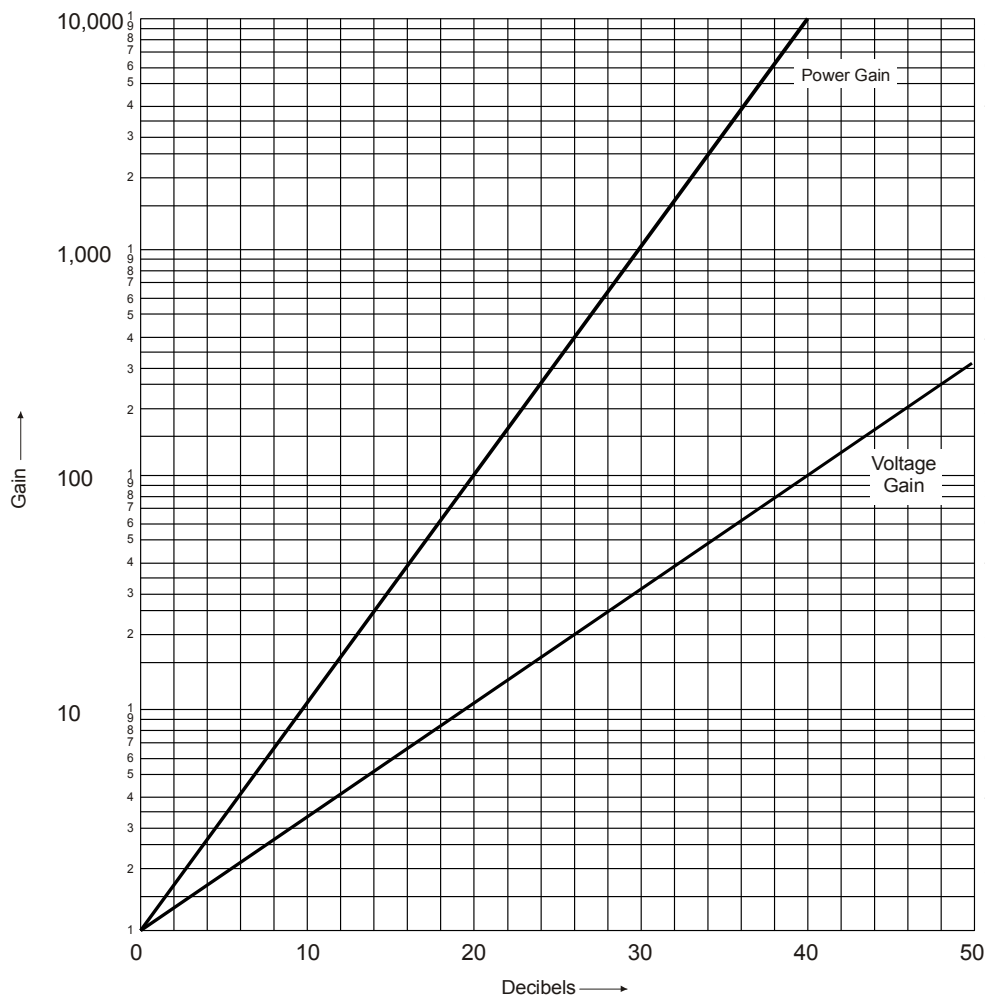
$$\text{Power gain} = 20 \log \frac{I_{\text{out}}}{I_{\text{in}}} \text{ dB}$$

29. The voltage at the half-power points in Fig 10 compared with the normal voltage of the amplifier is  $\sqrt{\frac{1}{2}}$ , since power =  $\frac{V^2}{R}$ . This ratio is 0.707 and so the voltage at the half-power points is approximately 70% of the normal voltage.



30. Fig 11 is a graph relating decibels to voltage and power ratios.

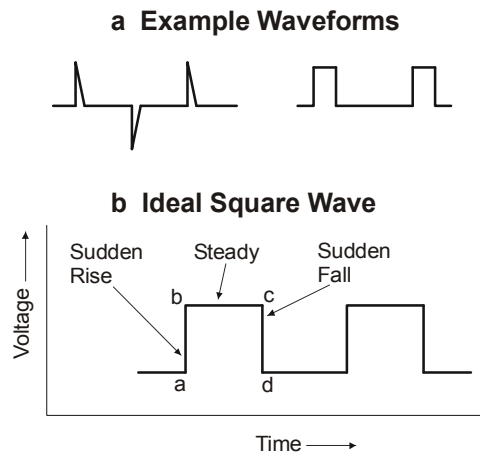
**14-14 Fig 11 Conversion to Decibels from Voltage and Power Ratios**



### Video Amplifiers

31. Audio frequency signals are electrical representations of waveforms which can be heard. In radar and television, we need to amplify electrical signals which correspond to waveforms which can be seen. These signals are known as video signals.

32. Some video waveforms are shown in Fig 12. The important point to note is that some parts of the waveform consist of sudden changes in voltage and other parts are formed by steady or slowly changing voltages. The ideal square wave shown in Fig 12 is a typical video waveform. It rises to its maximum voltage in a very short time, ideally in zero time, as shown by points a and b. Then it remains at a steady voltage (b to c) and at c falls sharply to its original value (point d). The steady parts indicate a very low frequency; the sudden changes indicate high frequencies. In fact, the square wave is made up of a low fundamental frequency plus many harmonics, or multiples, of this frequency. If we want to amplify the square wave without distorting it, we must amplify all these frequencies by the same amount. This means designing an amplifier which can give equal amplification to a range of frequencies; in theory the range of frequencies is infinite but, in practice, it must cover several MHz. For TV signals, the range is about 30 Hz to 4.5 MHz. This is the major difference between the video amplifier and the audio amplifier.

**14-14 Fig 12 Ideal Video Wave-forms**

### Radio Frequency (RF) Amplifiers

33. Amplifiers designed to handle signals from 20 kHz up to many thousand megahertz are called radio frequency amplifiers. They are used in transmitters to amplify the input signal from the receiver aerial. The main difference between RF amplifiers and the amplifiers so far considered is that one particular RF amplifier handles only a very narrow band of frequencies.

34. When a receiver is tuned to a particular radio station the RF amplifier is adjusted so that it amplifies only a narrow band of frequencies; signals from other stations on different frequencies are rejected. If another station is required the RF amplifier is adjusted to respond to the new station's carrier frequency. Therefore, although the RF amplifier only accepts a narrow band of frequencies at any one time, the band can be chosen from a wide range of frequencies by changing the value of the tuning capacitor.

35. The same principles apply to both television and VHF broadcasting. A TV station may have a carrier frequency of 45 MHz and the signals occupy the band of frequency 42.5 MHz to 47.5 MHz. In this case the band is wider than that for sound broadcasting and it is different again at VHF; but in all types of transmission there is a radio frequency carrier wave with a narrow band of frequencies on either side.

36. The RF voltage induced in the receiving aerial is in the order of microvolts and it must be amplified many times before any use can be made of it. Usually some amplification is carried out at the RF signal frequency, some at a lower radio frequency and some at audio or video frequency. The amplification at the original signal frequency is done by the RF amplifier, but, in addition to amplifying the signal, this stage also selects the desired station. This process of selection is called tuning. Therefore when the tuning dial on a receiver is turned the frequency to which the RF amplifier stage responds is altered.

37. The amplification is done by the transistor and the selection of the narrow band of frequencies to be amplified is performed by the tuned circuits which form part of the amplifier circuit. The tuned circuits consist of coils and capacitors which are designed to resonate at the required frequency. Tuning is the process of bringing the tuned circuit to resonance.

### Direct-coupled Amplifiers

38. A direct-coupled amplifier is one in which the output of one stage is connected directly into the input of the next stage and not via a coupling capacitor or transformer. Direct-coupled amplifiers are used to amplify voltages which change in value at a very slow irregular rate, ie voltages at very low frequencies. They are used in radar equipments, in analogue computers and in power supply circuits.

## AMPLIFIERS WITH FEEDBACK

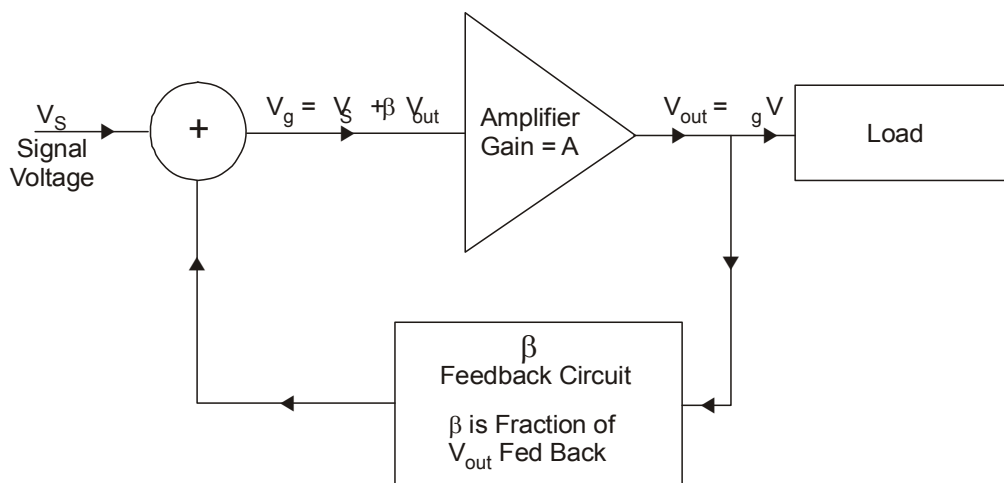
### Introduction

39. If a fraction of the output of an amplifying device or amplifying stage is added to its input, then feedback is said to occur. Should this feedback be in-phase with the input then it is classed as positive feedback and oscillation takes place, (this is the subject of Volume 14, Chapter 15). When a part of the output is fed back in antiphase, the feedback is said to be negative.

### Principle of Negative Feedback

40. Fig 13 depicts an amplifier with an “open-loop” gain ( $A$ ), a signal voltage ( $V_s$ ) and a feedback fraction of  $\beta$ .

**14-14 Fig 13 Block Diagram Representation of a Feedback Amplifier**



41. The 'closed-loop' gain, ie the gain with feedback, can be expressed by the following equations:

$$V_{out} = AV_g = A(V_s + \beta V_{out}) \quad \dots\dots (in\ general)$$

For negative feedback,  $\beta$  is negative,

$$\text{Hence, } V_{out} = A(V_s - \beta V_{out})$$

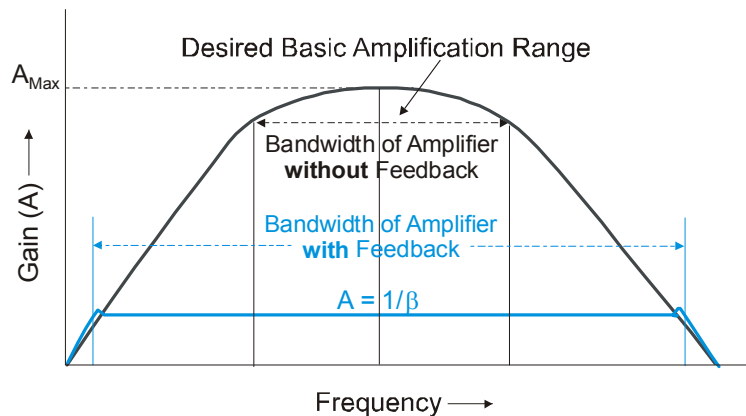
$$\text{and, } \frac{V_{out}}{V_s} = \frac{A}{1 + \beta A}$$

42. If the 'open-loop' gain of the amplifier is very high then the effective gain, with feedback, closely approaches  $1/\beta$ . The gain of the feedback amplifier is therefore independent of the gain of the basic amplifier. Variations in power supplies, ageing of components and other causes of variation in basic gain, have little effect on an amplifier operating with negative feedback.

### The Effect of Feedback on Frequency Response

43. The upper curve of Fig 14 represents the frequency response of a basic amplifier; that is to say the device's ability to amplify over a specified range of frequencies. Ideally the frequency curve should be flat over the desired range in order to provide uniform amplification.

**14-14 Fig 14 Frequency Response Curves of an Amplifier with and without Negative Feedback**



In practice this does not happen, mainly due to the frequency sensitivity of the passive components associated with the amplifying device, eg capacitors and inductors.

44. The lower curve represents the frequency response of the same amplifier, but with negative feedback applied. The result is a flatter response curve with a central gain of  $1/\beta$ . The fall-off in gain at the top and bottom ends of the band does not occur until the basic amplifier gain ( $A$ ) has fallen to such a level as to become insignificant in the feedback equation.

### Frequency Selective Feedback

45. The type of feedback discussed so far assumes a constant value of  $\beta$  regardless of the frequency of the input signal; this can be achieved using a simple resistive network. However, there are occasions when, for example, frequencies at the lower end of the band need to be amplified more than the ones at the top end, a form of bass boost. By introducing capacitance into the feedback network  $\beta$  becomes a variable value. High frequencies result in low reactance and a good deal of feedback, while low frequencies result in high reactance and reduced feedback.

### Input and Output Impedances with Feedback

46. The application of negative feedback has a marked effect on the input and output impedances of an amplifier, and this is summarized by Table 2. These changes with respect to an amplifier without feedback depend on the method of feedback tapping and injection. Feedback tapping can take the form of voltage or current, and injection can either be series or parallel.

**Table 2 Changes to Input and Output Impedances due to Feedback**

	Series	Parallel	
<b>Voltage</b>	Higher	Lower	Input Impedance
	Lower	Lower	Output Impedance
<b>Current</b>	Higher	Lower	Input Impedance
	Higher	Higher	Output Impedance

47. These changes in impedance values may be used to some advantage when matching the output of one stage to the input of another. The emitter follower (Fig 4) is a good example, where the application of series voltage feedback makes the device ideally suited for matching a video output amplifier to a transmission line, or coaxial cable.

## **POWER AMPLIFIERS**

### **Introduction**

48. The amplifiers considered so far have, in the main, been designed to deal with voltage amplification where the reduction of distortion has been a major consideration. Although such amplifiers develop power in their loads, this power is of little importance. Power amplifiers, however, are those in which the power is of chief consideration.

49. Voltage amplifiers normally operate over the linear part of the device's characteristic and with comparatively small voltage swings, well within the supply voltages. On the other hand, power amplifiers must use all the voltage available in order to operate efficiently.

### **Classes of Operation**

50. Voltage amplifiers normally work with a circuit arrangement known as Class A. Under these conditions the amplifier is regarded as being linear and provides low distortion but with low efficiency. The output current flows over the whole of the input cycle.

51. By biasing the input to the amplifier so that output current flows for only half the input cycle, greater efficiencies can be achieved. In order to duplicate the input waveform successfully, two devices working in a push-pull arrangement are required. This class of operation is known as Class B. With push-pull arrangement, Class AB can be used in order to reduce crossover distortion.

52. Radio frequency (RF) power amplifiers normally work under Class C conditions where the output current flows for less than half the input cycle. Although more efficient than other types of amplifier this class of amplifier introduces more distortion.

## CHAPTER 15 - OSCILLATORS

Contents	Page
Introduction .....	1
Maintenance of Oscillation .....	2
Frequency Stability of Oscillators .....	3
Crystal-controlled Oscillators .....	3
Frequency Multiplication .....	4
Frequency Synthesizers .....	4
Relaxation Oscillators .....	5
Microwave Oscillators .....	5

### Figure of Tables

14-15 Fig 1 A Simple Oscillatory Circuit .....	2
14-15 Fig 2 An Oscillatory Circuit .....	2
14-15 Fig 3 Crystal Used for Frequency Control .....	4
14-15 Fig 4 Single Crystal Frequency Synthesizer .....	4

### Introduction

1. An oscillator is a circuit, containing active and passive components, which converts DC power into AC power at a frequency determined by the values of the components. In many cases, the output waveform is sinusoidal and the oscillator is then referred to as a sine-wave or harmonic generator. Oscillators which generate waveforms such as sawtooth, square-wave and triangular, are known as relaxation oscillators.

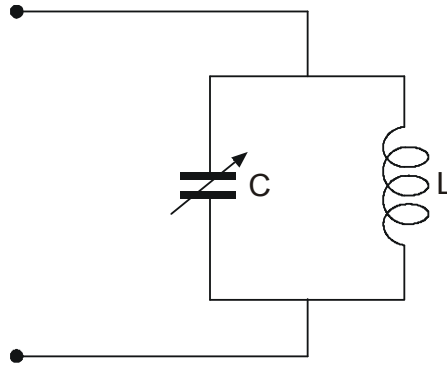
2. The main uses of oscillators are:

- a. The generation of radio frequency (RF) for transmission.
- b. The generation of RF for test equipment.
- c. The generation of an intermediate frequency (IF) in superheterodyne receivers
- d. The production of timing pulses in radar transmitters.

### Maintenance of Oscillation

3. If an electrical impulse is applied to the circuit shown in Fig 1, the circuit will oscillate or 'ring' at its natural frequency. However, the resistance in the circuit gradually uses up the power contained in the original impulse and the amplitude of the oscillation decays exponentially. This is known as damping. (The process of natural oscillation is discussed in detail in Volume 14, Chapter 4.)

**14-15 Fig 1 A Simple Oscillatory Circuit**

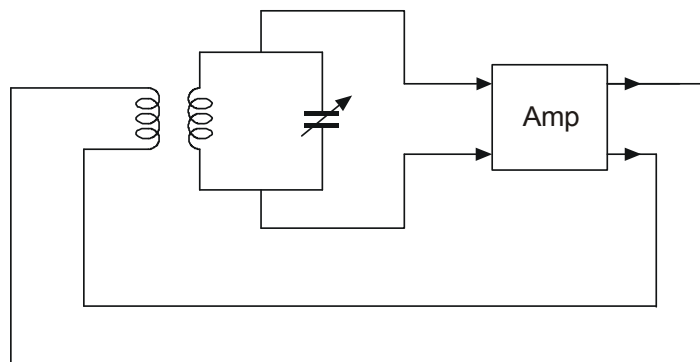


4. If, on the other hand, the circuit can be continually supplied with power at the right frequency and in the correct phase, the oscillations may be maintained at a constant level. This can be achieved by applying the oscillations to the input of an amplifier and then feeding part of the output back to the input, in phase. This is shown diagrammatically in Fig 2. Such an arrangement is called an oscillator, and the feedback required for the maintenance of oscillation is 'positive'. The frequency of oscillation depends on the values of the inductance and the capacitance in the LC 'tank' circuit. It is given, to a very close approximation, by the equation:

$$f_o = \frac{1}{2\pi\sqrt{LC}} \text{ Hz}$$

where L is in henries and C is in farads. This equation shows that the resonant frequency,  $f_o$ , increases with a decrease in either L or C. If the feedback is more than enough to compensate for the damping effect, it might be expected that the oscillations would build up to an infinite amplitude. In fact, the oscillations reach a maximum value determined by the flattening of the amplifier's characteristics.

**14-15 Fig 2 An Oscillatory Circuit**



## Frequency Stability of Oscillators

5. One of the most important features of an oscillator is its frequency stability, ie its ability to provide a constant frequency output under varying conditions. Some of the factors which affect the frequency stability of an oscillator, and some of the methods used to counteract frequency drift, are as follows:

- a. **Temperature.** Coils and capacitors alter in size, and therefore in value, with temperature changes. This can be compensated for by enclosing the tuned circuit in a temperature controlled compartment, or by using a capacitor whose value decreases with an increase in temperature, thus counteracting the corresponding increase in value of the inductor.
- b. **Variations of Load.** The output from an oscillator is coupled to some other device, eg an aerial or an amplifier. Any change in the value of this load causes the oscillator frequency to drift. This effect can be reduced by taking only a fraction of the power which the oscillator could provide, ie loosely coupling the oscillator to its load. Alternatively, a buffer amplifier can be placed between the oscillator and load. This buffer stage screens the oscillator circuit from variations in load impedance.
- c. **Changes in Power Supplies.** If the power supplies to the oscillator circuit vary, the output frequency will change. To avoid this, power supplies are carefully stabilized.
- d. **Vibration and Shock.** If mobile equipment is subject to vibration, it can cause changes in oscillator frequency. Frequency drift due to this cause is reduced by mounting the equipment on shock absorbers.
- e. **Hand Capacitance.** The proximity of the operator's hand or other parts of the body may introduce extra capacitance into the oscillatory circuit. This can be reduced by earthing one side of the tuned circuit or by screening the whole circuit.

## Crystal-controlled Oscillators

6. A class of oscillator which has exceptionally good frequency stability is the crystal-controlled oscillator. It is widely used when oscillations at one fixed frequency are required, but it cannot be tuned over a range of frequencies as can the LC oscillator. However, by using harmonics generated by several crystal oscillators, a number of spot frequencies in a wide range can be chosen.

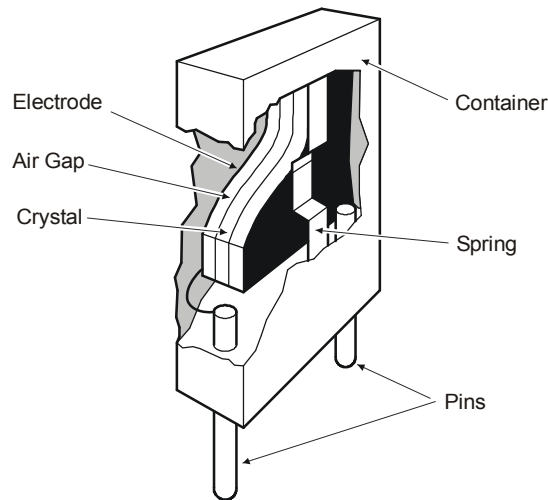
7. If a voltage at the same frequency as the natural frequency of the crystal is applied to opposite faces of the crystal, resonance occurs. The crystal's natural frequency is very stable and will remain constant to within one part in  $10^6$ . If temperature compensation and other arrangements are included, stability of 1 in  $10^7$  can be obtained. The range of frequencies covered by normal crystals is from 50 kHz to 9 MHz.

8. The crystal which controls the frequency is the same type as that used in the crystal microphone; it is made from Rochelle salt or quartz. These crystals exhibit a piezo-electric effect, ie they develop a voltage across opposite faces when they are compressed or expanded and they contract and expand (ie oscillate) when an alternating voltage is applied across them. Such a crystal has a natural frequency of oscillation which depends on its size, its thickness, and the way in which it has been cut; a thick crystal will oscillate at a lower frequency than a thin one.



9. The construction of an oscillator control crystal is shown in Fig 3. A thin crystal wafer is mounted between two metal plates from which electrical connections are taken.

**14-15 Fig 3 Crystal Used for Frequency Control**



### Frequency Multiplication

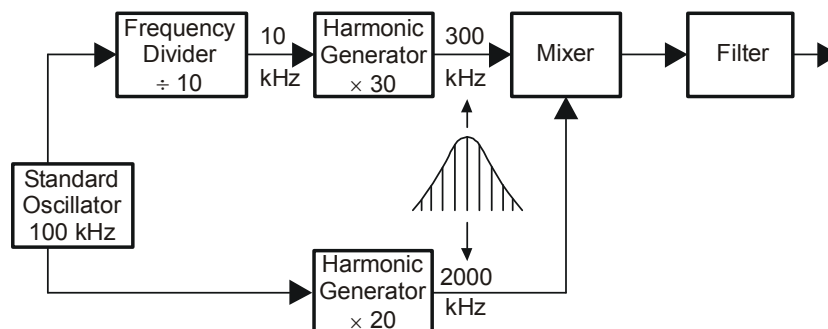
10. For frequencies above 14 MHz, it is not practical for quartz crystals to be ground thinly enough for general usage. If a crystal is used at a relatively low frequency, it can be used to control a non-linear amplifier which generates harmonic multiples of the fundamental. The crystal oscillator is more stable at a lower frequency.

11. If the LC output circuit of an amplifier is tuned to a harmonic of the input then the circuit receives a 'kick' every second or third cycle of its oscillation, which is sufficient to keep it going. In this way, any frequency may be multiplied up as many times as are desired, although it is not usual to multiply by factors greater than three in one stage.

### Frequency Synthesizers

12. A frequency synthesizer is basically a circuit in which harmonics and sub-harmonics of a single standard oscillator are combined to provide a multiplicity of output signals which are all harmonically related to a sub-harmonic of the standard oscillator. A simple block diagram is shown in Fig 4.

**14-15 Fig 4 Single Crystal Frequency Synthesizer**



13. The output from the  $\times 30$  harmonic generator is a spectrum of frequencies, centred on 300 kHz, with a separation of 10 kHz. Similarly, the output from the  $\times 20$  harmonic generator is a spectrum of frequencies, centred on 2,000 kHz, with a separation of 100 kHz. The desired 'harmonic mix' is selected by the filter. A great advantage of this circuit is that accuracy and stability of the output signal is essentially equal to that of the standard oscillator, which can be tightly controlled.

### **Relaxation Oscillators**

14. Relaxation oscillators generate non-sinusoidal outputs, the most important examples of these being the rectangular or square wave, and the sawtooth wave. Square waves are of immense importance in digital computers and radar, whilst sawtooth generators are used extensively as a source for the time-base waveforms associated with cathode-ray tubes.

15. The most commonly used square wave oscillator is the multivibrator. This generator, in its basic form, consists of two RC cross-coupled amplifiers; the output of one being connected to the input of the other, and vice versa. Normally, one device is conducting while the other is cut-off, and the change-over is determined by the RC time constant. The multivibrator action ensures rapid switching from the 'off' to the 'on' IF state for each amplifier, resulting in a square-wave output with sharp leading and trailing edges.

### **Microwave Oscillators**

16. At frequencies higher than about 1,000 MHz, conventional low frequency devices and resonant circuits become inefficient. Most microwave components and equipment take on a different form. For example, a resistor is replaced by an attenuator, an LC tuned circuit by a resonator, a connecting wire or cable by a waveguide. The cavity magnetron and the reflex klystron are well-established examples of microwave oscillators. These devices, along with other methods of microwave generation, are discussed in Volume 14 Chapter 20.

## CHAPTER 16 - TRANSMITTERS

Contents	Page
Introduction .....	1
KEYING AND MODULATION .....	3
General .....	3
Transmitter Keying .....	3
Analogue Modulation.....	3
TRANSMITTER SYSTEMS.....	6
Amplitude Modulated Transmitters.....	6
Frequency Modulated Transmitters .....	7
Pulse Modulated Transmitters.....	7

### Table of Figures

14-16 Fig 1 The MOPA Transmitter .....	2
14-16 Fig 2 A Constant Frequency, Constant Amplitude, RF Wave.....	2
14-16 Fig 3 Transmitter Keying - Morse Code .....	3
14-16 Fig 4 Amplitude Modulation.....	4
14-16 Fig 5 Frequency Modulation.....	5
14-16 Fig 6 Methods of Pulse Modulation.....	6
14-16 Fig 7 High-level Modulation AM Transmitter.....	7
14-16 Fig 8 A Basic FM Transmitter.....	7
14-16 Fig 9 High Power Oscillator Transmitters .....	8
14-16 Fig 10 A MOPA Pulse Transmitter.....	8

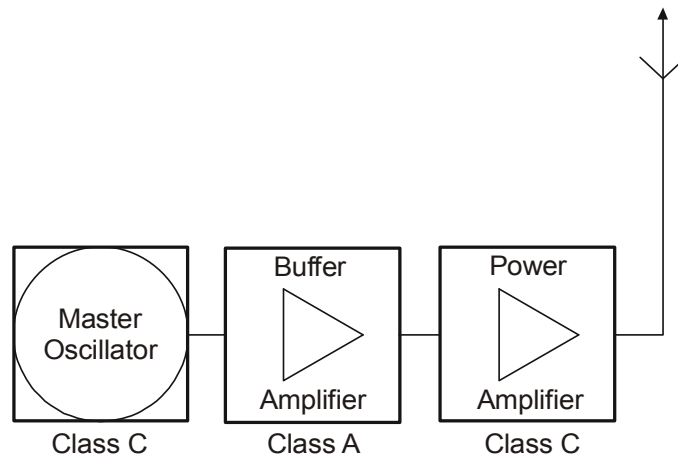
### Introduction

1. Many types of practical transmitters are in use, ranging from low frequency ground installations of considerable size, weight, and power, to miniaturized VHF and UHF equipment for airborne or man-portable applications. Although the appearance and uses of these transmitters may vary considerably, they all use the same basic principles.

2. The simplest form of transmitter comprises an oscillator connected to an aerial. However, this simple arrangement produces only a limited amount of radiated power, and suffers from poor frequency stability, since any oscillator from which more than a very little power is drawn tends to drift in frequency.

3. In a practical transmitter, these drawbacks are overcome by incorporating amplifiers between the oscillator and the aerial. The arrangement is illustrated in Fig 1 and is known as the Master Oscillator Power Amplifier (MOPA) System. The main components of the system and their functions are:

- a. **Master Oscillator.** The function of the master oscillator is to produce a radio frequency voltage of good frequency stability. A crystal-controlled oscillator is normally used. To enhance frequency stability, it is normally run at low power and may have its crystal mounted in a thermostatically temperature-controlled enclosure.

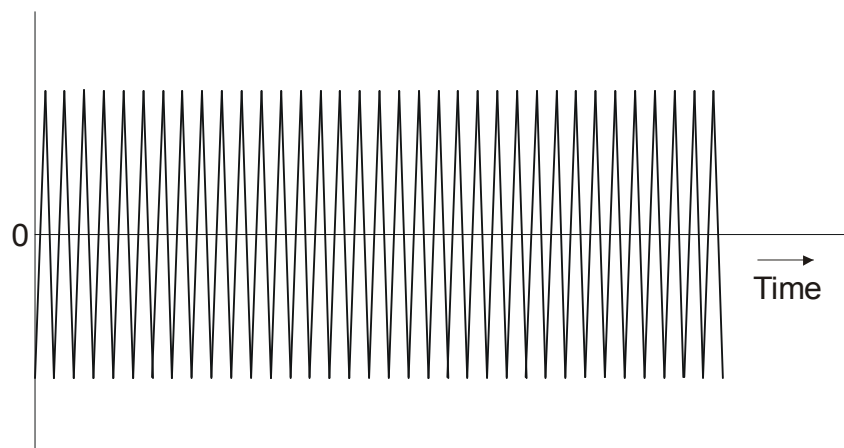
**14-16 Fig 1 The MOPA Transmitter**

b. **Buffer Amplifier.** The buffer amplifier provides a constant light load for the master oscillator, thereby isolating the oscillator from load variations, which improves frequency stability. In addition, the buffer amplifier amplifies the RF signal from the oscillator to provide the necessary drive to the power amplifier. The buffer amplifier normally operates under Class A conditions (see Volume 14, Chapter 14, Para 50).

c. **Power Amplifier.** The function of the RF power amplifier is to convert the RF drive from the buffer amplifier to a sufficiently high power level to feed the aerial and provide adequate energy for radiation. The power amplifier operates at a high power and at high efficiency under Class B or C conditions.

4. In many instances, especially at VHF or above, the frequency of the output at the aerial is required to be higher than that at which the master oscillator can efficiently maintain a stable frequency. In this situation, frequency-multiplier stages (doublers or treblers) are inserted between the master oscillator and the final power amplifier stage.

5. The simple MOPA transmitter described would transmit a constant frequency, constant amplitude, RF wave (Fig 2), ie it would carry no information. In order to transmit information, the transmitter output must be altered in some way.

**14-16 Fig 2 A Constant Frequency, Constant Amplitude, RF Wave**

## KEYING AND MODULATION

### General

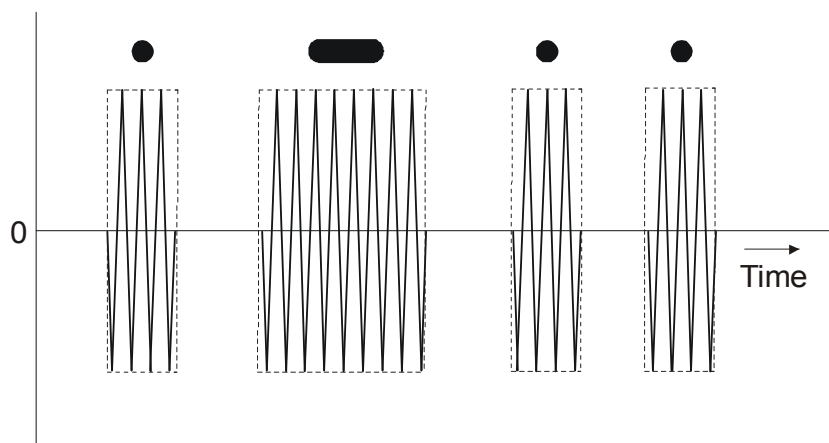
6. The techniques of superimposing information onto a transmitted signal are covered in Volume 14, Chapter 22. However, for convenience, the basic ideas will be reviewed in the following paragraphs, since the technique employed affects the arrangement of a practical transmitter.

7. The information to be transmitted may be either in the form of a quantity which varies continuously with time, e.g. speech, which is termed an analogue signal, or it may be in a form which is only permitted to have discrete values or levels, e.g. Morse code, which is known as a digital signal.

### Transmitter Keying

8. In order to transmit Morse (or any similar) code by the continuous wave of Fig 2, it is necessary to switch the transmitter on and off in such a way that radiation from the aerial occurs for short and long periods of time corresponding to the dots and dashes of the code (Fig 3) - the operation is known as 'keying'. One method of interrupting the transmitted output would be to switch the master oscillator on and off, but this is not normally done since it may cause the frequency to drift. Instead, the switching is normally applied to one of the other stages in the transmitter.

14-16 Fig 3 Transmitter Keying - Morse Code



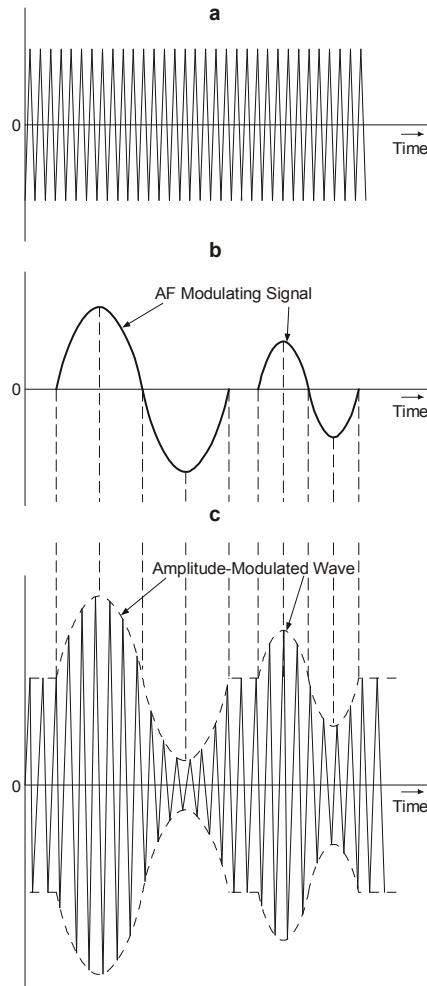
### Analogue Modulation

9. Analogue information is transmitted by changing the amplitude, frequency, or phase (which has limited applications and is not considered further here) of the output (carrier) signal, to reflect changes in the information signal. Alternatively, the transmitter radiation may be concentrated into short bursts of RF energy, whose amplitude, length, or interval can be varied (pulse modulation).

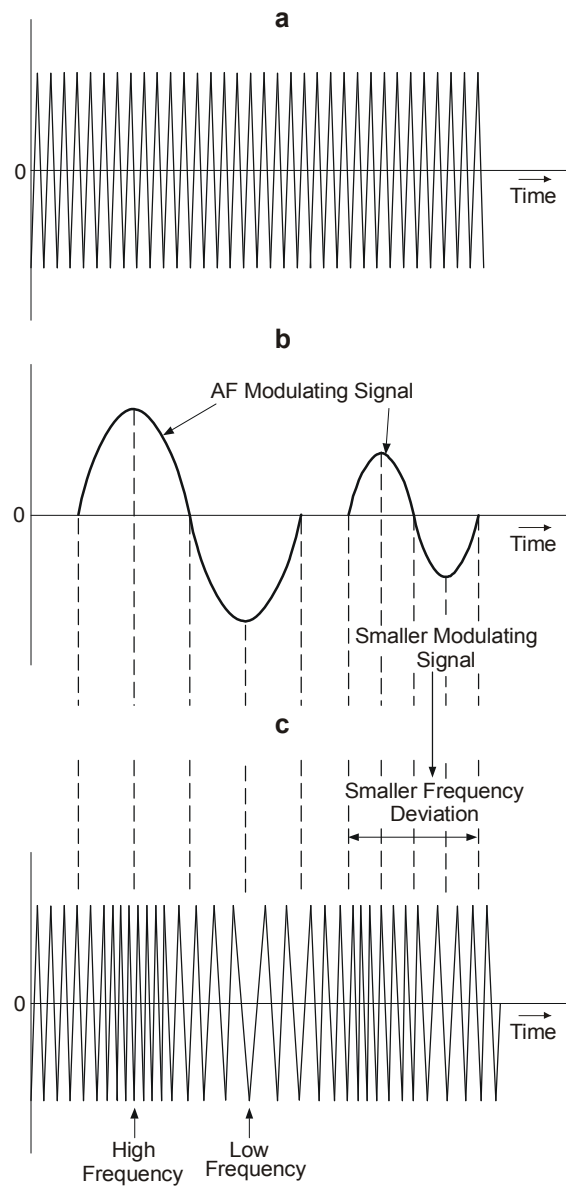
10. **Amplitude Modulation (AM).** In amplitude modulation, the modulating signal is used to modify the amplitude of the RF carrier signal without affecting its frequency (Fig 4). Fig 4a represents the unmodulated CW carrier signal from the transmitter. Fig 4b shows the audio frequency modulating signal, and Fig 4c shows the change in the amplitude of the carrier, above and below the unmodulated

level, proportional to the amplitude and sign of the modulating signal. The rate of change of amplitude of the carrier depends on the modulating frequency. The outline of the modulated wave, known as the modulation envelope, is an exact replica of the modulating signal. Single sideband systems (SSB) are a variation of AM.

**14-16 Fig 4 Amplitude Modulation**

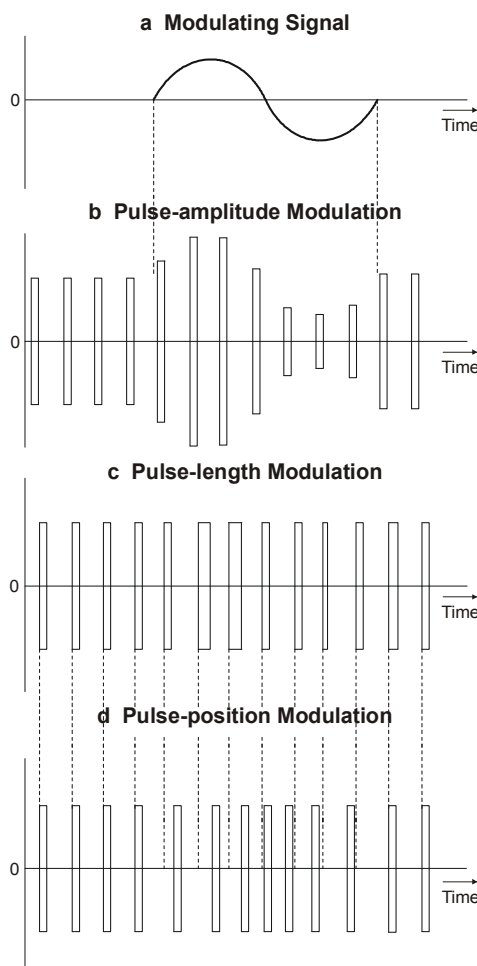


11. **Frequency Modulation (FM).** In frequency modulation, the modulating signal is used to modify the carrier frequency without affecting its amplitude (Fig 5). Fig 5a represents the unmodulated CW carrier signal from the transmitter. Fig 5b shows the audio frequency modulating signal and Fig 5c shows the changes in the frequency of the carrier, above and below the unmodulated frequency, proportional to the amplitude and sign of the modulating signal. The rate of change of carrier frequency depends on the modulating frequency.

**14-16 Fig 5 Frequency Modulation**

12. **Pulse Modulation (PM).** Pulse modulation can be used to convey information by varying the pulse amplitude (Fig 6b), the pulse length (Fig 6c), or the interval between pulses (Fig 6d) in accordance with the modulating signal (Fig 6a).

**14-16 Fig 6 Methods of Pulse Modulation**

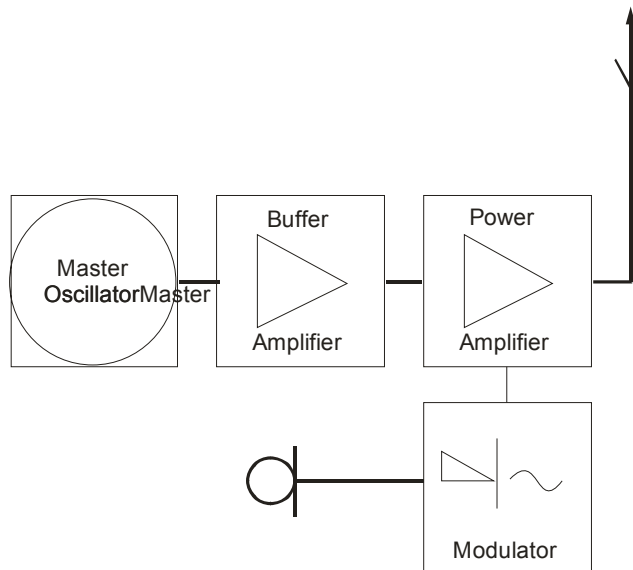


## TRANSMITTER SYSTEMS

### Amplitude Modulated Transmitters

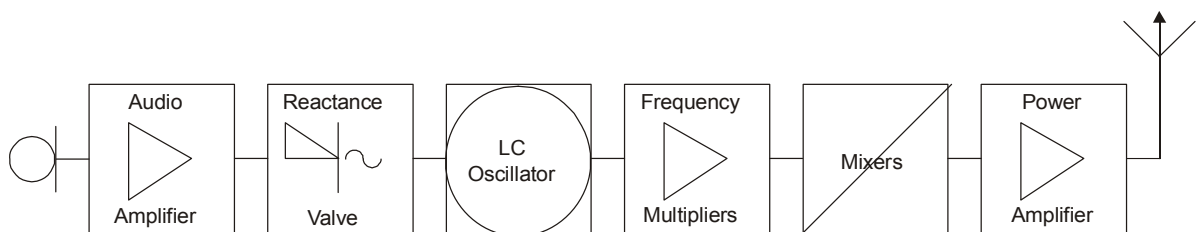
13. The modulating signal in an AM system may be a keyed audio note representing the dots and dashes of the Morse code, or it may be an analogue signal derived, for example, from a microphone. Modulation is said to be high or low, depending on whether it is applied in the transmitter at a point of high or low RF power. If the modulation is applied at an early stage, ie low-level modulation, then the following stages must be operated in Class A or B in order to avoid distorting the modulated RF input. A block diagram of a high-level system is shown in Fig 7. The RF carrier is generated by the oscillator and then amplified to the power level required for radiation from the aerial. The AF amplifiers raise the power of the AF signal from the microphone to the level required for modulation. The AF signal amplitude modulates the RF carrier in the RF power amplifier stage. For a given total output power, low-level modulation would need very little modulation or audio frequency power, but a large pre-amplifier (PA) stage. The PA stage would be inefficient since it would have to be operated in Class A or B, but offsetting the disadvantage, the AF stage would be small without the need for a heavy, high power, AF transformer. High-level modulation uses an efficient Class C amplifier at the PA stage, but uses a bulky, high power, inefficient AF amplifier (the modulator).



**14-16 Fig 7 High-level Modulation AM Transmitter**

### Frequency Modulated Transmitters

14. A block diagram of a basic FM transmitter is shown in Fig 8. The AF signal is amplified and passed to the reactance device, which is used to vary the frequency produced by the oscillator. The reactance modulator will produce only small frequency deviations, and a practical FM transmitter must raise the frequency deviation by using frequency multipliers. For example, if the output from the oscillator is 2 MHz + 650 Hz, it is necessary to convert this to a VHF output of say 90 MHz with a frequency deviation of 75 kHz. To change the frequency deviation from 650 Hz to 75 kHz requires multiplication by about 115. Since this would produce a carrier of 230 MHz, which is too high, means must be found of bringing the carrier back to 90 MHz without affecting the frequency deviation. This is achieved by the mixer, which would be fed by an RF oscillator at a frequency of 140 MHz (230 MHz minus 90 MHz).

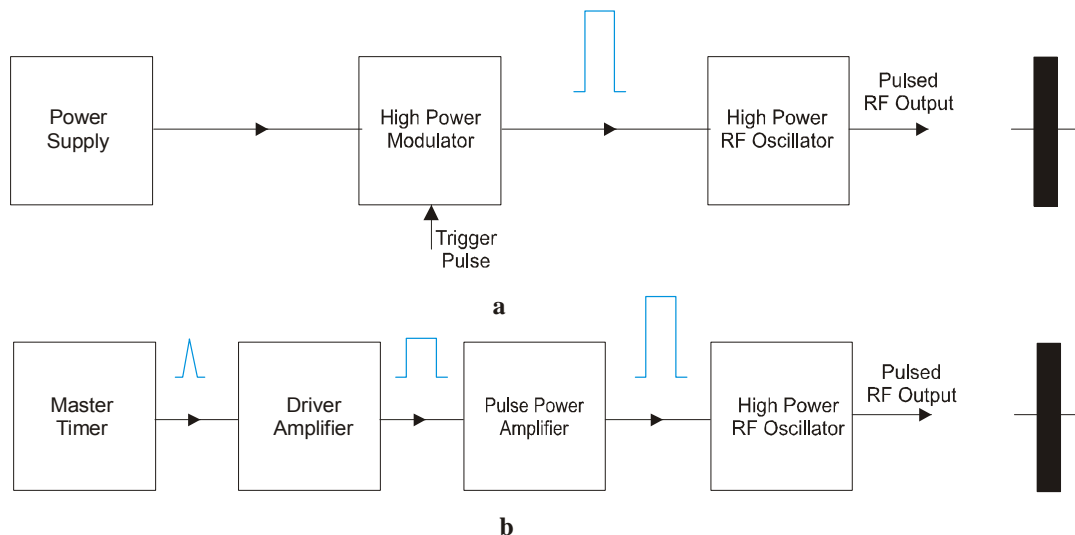
**14-16 Fig 8 A Basic FM Transmitter**

### Pulse Modulated Transmitters

15. A pulse-modulated transmitter may be either a higher power oscillator type or a MOPA type. The choice between the two configurations is governed mainly by the application. Transmitters that utilize power oscillators are usually smaller than MOPA transmitters, but the latter are more stable and are usually capable of providing greater mean power. Power oscillators are therefore likely to be found in applications where small size and portability are of the greatest importance, and MOPA transmitters in radar applications in which high power or good MTI performance are desired.

16. **The High Power Oscillator Transmitter.** The high power oscillator, typically a magnetron (see Volume 14, Chapter 20), is switched on and off by the modulator. The modulator may be either a high power stage, as shown in Fig 9a, or it may consist of several stages which amplify and shape the master timing pulse, as shown in Fig 9b. In both cases, the RF energy of the pulse is generated in the oscillator and forms the final RF output.

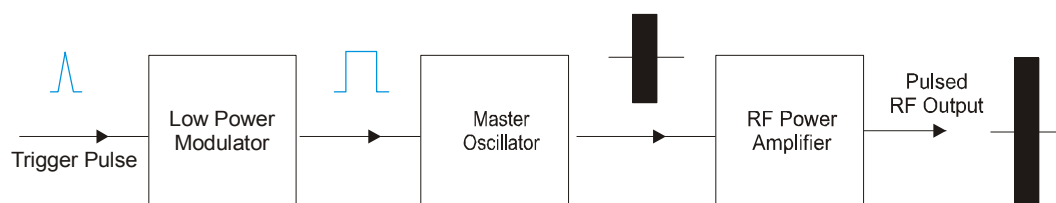
14-16 Fig 9 High Power Oscillator Transmitters



17. **The Master Oscillator Power Amplifier (MOPA).** In the MOPA type of pulse transmitter (Fig 10), a low power RF oscillator is amplitude modulated, and its output is then amplified by the RF amplifier. The master oscillator may be crystal controlled, or it may be a stable frequency resonant cavity oscillator. The MOPA system has the following advantages over other systems:

- The frequency stability is much better.
- A simple modulator can be used because it does not have to provide the high power required to feed a single stage transmitter.
- Phase coherency between radiated pulses is much easier to obtain.

14-16 Fig 10 A MOPA Pulse Transmitter



## CHAPTER 17 - RECEIVERS

Contents	Page
Introduction .....	1
Selectivity and Choice of IF .....	3
COMMUNICATION RECEIVERS .....	4
General .....	4
Single Sideband (SSB) Receivers .....	5
FM Receivers .....	5
Frequency Shift Keying (FSK) Receivers .....	6
RADAR RECEIVERS .....	6
Introduction .....	6
CW Receivers .....	6
Pulse Doppler Receivers .....	7
Logarithmic Receivers .....	8

### Figures

14-17 Fig 1 A Basic Radio Frequency Receiver .....	1
14-17 Fig 2 A Superheterodyne Receiver .....	2

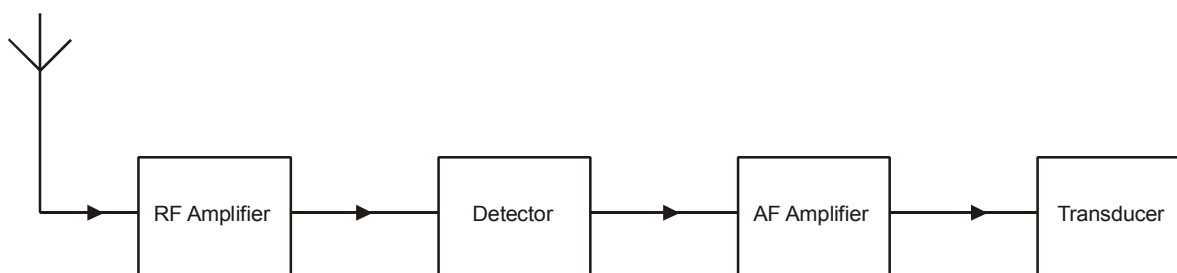
### Table

Table 1 Typical Intermediate Frequencies .....	4
------------------------------------------------	---

### Introduction

1. The task of a radio frequency receiver is to intercept some of the RF energy radiated from the transmitter, to detect the information it contains, and to reproduce it in an acceptable form. The basic receiver shown in block form in Fig 1 uses an RF amplifier to select and amplify the required frequency, a detector to extract the information, and an audio frequency amplifier to produce enough power to operate the loudspeaker (transducer). This type of receiver could be used for single channel operation, but has limited potential for tuning to other channels.

**14-17 Fig 1 A Basic Radio Frequency Receiver**

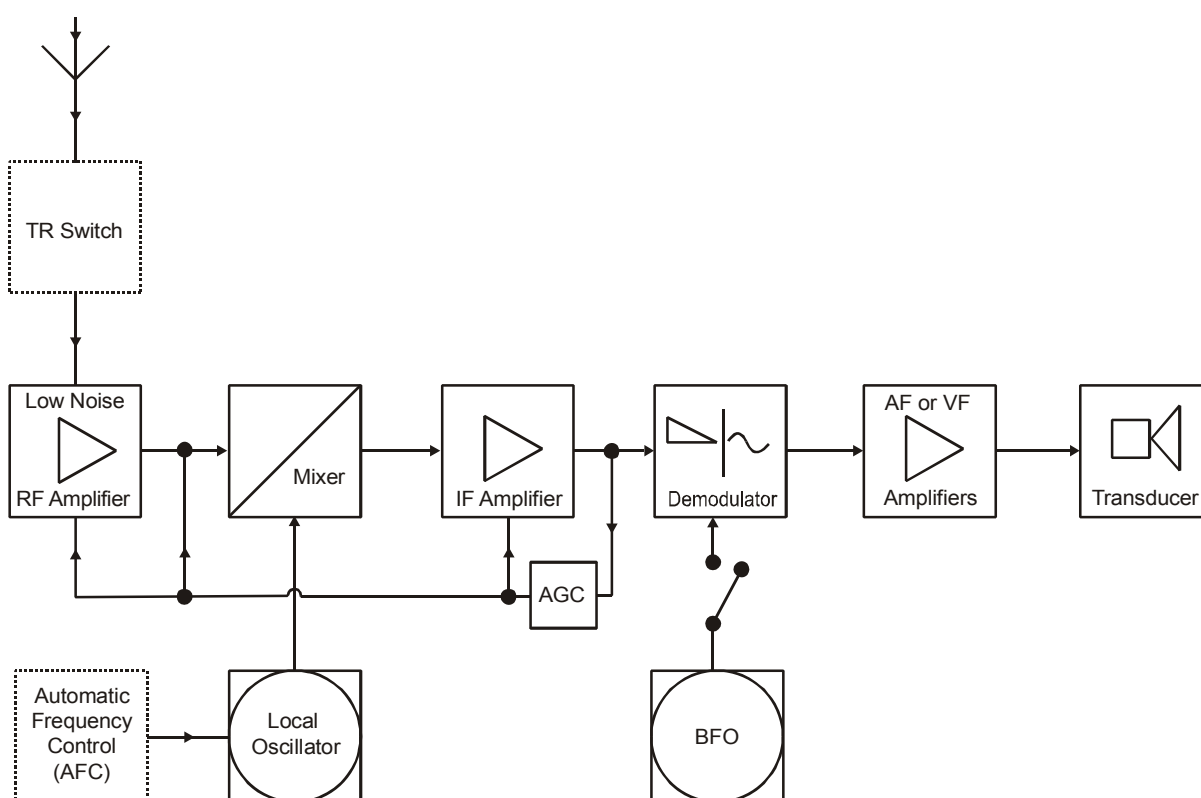


2. A practical receiver uses the superheterodyne (often abbreviated to superhet) principle in which all incoming signals are changed to a fixed frequency (the intermediate frequency - IF) for amplification

before demodulation. Most of the amplification within the receiver takes place in the IF stages. Fig 2 depicts a multi-purpose, superhet receiver. The various stages are:

- a. **Transmit/Receive (TR) Switch.** The TR switch is used to allow the aerial to act as a radiator or receiving aerial. TR switches are discussed in Volume 14, Chapter 20.
- b. **RF Amplifier.** Although not always essential, an RF amplifier is usually included and is an RF voltage amplifier with Class A bias and a tuned circuit collector load. Its functions are to select the wanted signal from a background of other signals and noise, to give some pre-mixer amplification to the signal and hence improve the signal-to-noise ratio at the receiver output, and to isolate the aerial from the local oscillator so as to reduce radiation at the local oscillator frequency.

**14-17 Fig 2 A Superheterodyne Receiver**



- c. **Local Oscillator.** The local oscillator can be any one of the basic oscillator circuits provided that it is stable over the required frequency range. The local oscillator and the RF amplifier tuned circuits are 'ganged' together so that there is always a constant fixed difference between their frequencies. At HF and below, the local oscillator frequency is usually chosen to be higher than that of the incoming signal in order to receive signals of lower frequency than the IF, and to keep the ratio of the maximum and minimum frequencies within the range of a normal tuning capacitor. At VHF and above, the local oscillator operates below the signal frequency in order to improve stability. At these frequencies the ratio of the maximum and minimum frequencies is much lower and well within range of the tuning capacitor.

- d. **Mixer (First Detector).** The inputs to the mixer are the modulated signal and the local oscillator output, the mixer producing the sum and difference of these frequencies. The difference frequency is selected by a tuned circuit and this forms the IF signal which is modulated with the same waveform as the original RF signal. The choice of mixer circuit depends primarily

on the frequency of the input signal. At HF and below the multiplicative type is used, but above these frequencies additive mixing is used since this results in less interaction between the two inputs to the mixer. Interaction can be reduced further by the insertion of a buffer amplifier between the local oscillator and the mixer.

e. **IF Amplifier.** The IF amplifier is an RF amplifier operating at a low, fixed, radio frequency and because the frequency is fixed, this stage is always operating under optimum conditions. This permits the sensitivity, selectivity, and stability to be much higher than would be possible with a tuneable RF amplifier. In most receivers there are usually several IF amplifiers in cascade, forming what is known as the IF strip, and it is in this part of the receiver that most of the amplification is achieved. The IF amplifier also sets the bandwidth of the receiver overall.

f. **Demodulator (Second Detector).** The demodulator is invariably a diode detector since the amplitude of the signal at this stage in a receiver is sufficiently large (a few volts) to allow a diode to operate efficiently and with good linearity.

g. **AF or Video Frequency Amplifier.** The AF or video frequency amplifiers form the final processing stage in a receiver and are necessary to bring the signal to a level suitable for presentation. The number of amplifiers present in this stage is determined by the power output required.

h. **Beat Frequency Oscillator (BFO).** A beat frequency oscillator is required for the reception of CW signals. It produces a signal at a frequency which combines at the demodulator with the output of the IF stage to produce a component at the difference frequency, usually 1 kHz. This can now be detected and amplified to give an audio output. Since the BFO is not required when full AM signals are being received, an on/off switch is incorporated.

i. **Automatic Gain Control (AGC).** The automatic gain control circuit reduces the gain of the receiver in proportion to the input signal strength, and the output thus tends to remain constant despite signal fading and changes due to tuning. The circuit achieves this by feeding back a DC voltage (from a diode driven by the last IF stage) which is proportional to the signal strength at that point. This voltage varies the bias on the earlier stages, thus varying the gain of the controlled stages. A simple AGC circuit reduces the gain of the receiver for all signals, even weak ones. This is a disadvantage and so the normal arrangement is to use a delayed AGC circuit which does not feed back an AGC voltage until the signal rises above some predetermined value.

### Selectivity and Choice of IF

3. There are two main ways in which an unwanted signal may cause interference with the wanted signal in the superhet:

a. **Adjacent Channel Interference.** Adjacent channel interference is interference from a signal which is close in frequency to the wanted signal. It may be reduced or eliminated by making the tuned circuits of the IF amplifiers highly selective.

b. **Second (or Image) Channel Interference.** The carrier frequency of the wanted signal differs from the local oscillator frequency by the IF. However there is another frequency, also differing from the local oscillator frequency by the IF, but in the opposite sense to the wanted signal, which could cause interference. For example, consider a wanted carrier frequency of 2.2 MHz being received by a receiver with an IF of 500 kHz; the local oscillator would in this situation be tuned to 2.7 MHz. In addition to the wanted signal, a frequency of 3.2 MHz could also mix with the local oscillator to produce a 500 kHz difference frequency. This second channel interference can be prevented by ensuring that the image channel frequency lies well outside the pass band of the RF amplifier.

4. **Choice of IF.** The higher the IF, the further away is the second channel interference frequency, which makes it easier for the RF amplifier tuned circuit to discriminate between them. Conversely, a low value for the IF assists in discriminating between adjacent signals thereby reducing adjacent channel interference. The choice of IF must therefore be a compromise between these conflicting requirements. However, in high quality receivers where this compromise is not acceptable, the problem can be avoided by use of the double superhet receiver technique.

5. **The Double Superhet Receiver.** In the double superhet receiver the incoming signal is first changed to a high IF which assists in second channel rejection. After amplification at this frequency, a second mixer is used to produce a final low IF which improves adjacent channel rejection. Since the input to the second mixer has a fixed centre frequency, the second local oscillator can be preset and its tuning does not have to be ganged to the tuning capacitors of the first RF and first local oscillator stages.

6. **Typical Output.** Typical signal and intermediate frequencies for various types of input signal are shown in Table 1.

**Table 1 Typical Intermediate Frequencies**

Type of Signal	Signal Frequency	IF
AM Broadcast	1 MHz	454 kHz
FM Broadcast	90 MHz	10.7 MHz
Television	150 MHz	34 MHz
HF Communication	10 MHz	600 kHz
VHF/UHF Comms	300 MHz	15 MHz - 1st IF 2 MHz - 2nd IF

7. **Functional Differences.** Although most RF receivers use the superhet principle, there are differences between communications and radar receivers, and between types of receiver within each group. These characteristics will be reviewed in the following paragraphs.

## COMMUNICATION RECEIVERS

### General

8. The essential features of a communication receiver are:

- a. **Frequency Coverage.** No one receiver can possibly cover successfully the whole of the RF spectrum and so receivers are designed to operate in a particular band.
- b. **Sensitivity.** Sensitivity is a measure of the ability of the receiver to intercept weak signals and extract information from them and it depends upon the amount of RF amplification available at the beginning of the receiver. Sensitivity cannot be increased indefinitely by increasing the number of RF stages because of instability due to interaction and usually only one amplification stage is used.
- c. **Selectivity.** Selectivity is a measure of a receiver's ability to intercept only the required signal and to extract the information it carries, even though there may be other information carrying signals on close frequencies. An RF stage on its own is unable to provide the required

response for good selectivity and the superhet principle is almost universal since it is easier to obtain the required selectivity at lower frequencies.

d. **Fidelity.** Fidelity is a measure of how well the receiver reproduces the received baseband signal without distortion. Whereas effective sensitivity and selectivity require a narrow bandwidth, good fidelity can only be obtained with a wide bandwidth.

e. **Stability.** As a communication receiver can be remotely operated it is necessary for the local oscillator to have good frequency stability.

### Single Sideband (SSB) Receivers

9. The SSB receiver is a superhet arrangement but with the additional requirement of re-inserting the carrier at the correct frequency and amplitude. The carrier must be present in the receiver along with the sideband before demodulation can take place, but if the frequency is incorrect, the signal will be distorted, and if the amplitude is incorrect, the effective depth of modulation will be changed.

10. When the SSB signal is transmitted with a controlled carrier or a pilot carrier, a triple superhet receiver is often used. From the amplified composite signal, the carrier is selected by a filter, and used to produce an automatic frequency control (AFC) voltage (by means of a discriminator and reactance device) with which the frequency of the second local oscillator is controlled. If the transmitter drifts in frequency, the receiver will follow the drift because the pilot or controlled carrier indicates the direction and extent of the drift. A portion of the signal output from the third frequency changer is passed through a filter which rejects the sidebands and passes the carrier frequency only. After amplification and limiting, this is fed into the demodulator stage where it recombines with the sideband frequencies and enables demodulation to take place in the normal way.

11. With the suppressed carrier system there is no carrier present in the receiver to use as a reference, and so the frequency drift in both transmitter and receiver must be negligible. The oscillators are thus usually crystal and thermostatically controlled to provide this high standard of stability. The receiver is usually a double superhet, both local oscillators being controlled by the frequency of a very stable temperature controlled crystal oscillator which is also used to provide the re-insert carrier at the demodulation stage. The frequency at this point is that of the second IF; thus, the difference frequencies produced in the detector are the original audio frequencies, i.e. the modulation.

12. In airborne SSB systems, the receiver forms part of a combined transmitter-receiver, and the stable oscillators used in the frequency translation process in the transmitting function are also used as local oscillators in the receiver function. This helps to ensure the overall frequency stability of the system.

### FM Receivers

13. The FM receiver is a superhet arrangement but with wider bandwidth IF amplifiers compared to an AM receiver, and with a limiter and discriminator taking the place of the envelope detector in the demodulation stage. The limiter eliminates unwanted amplitude variations in a signal by cutting off the positive and negative extremities of the waveform. The discriminator is a device which produces a voltage proportional to the instantaneous deviation of the FM signal. This voltage will then be equivalent to the original modulating signal. An FM receiver, operating at VHF and above, is used for short-range communication and there is therefore usually a fairly strong input signal. AGC is normally only necessary in fringe areas and where it is employed the AGC diode would normally take its input from the last IF stage.

## Frequency Shift Keying (FSK) Receivers

14. The input to an FSK receiver is frequency modulated with very small deviation and so the basic requirements for the reception of FSK signals are a narrow bandwidth, stringent frequency stability, and selective filtering in the demodulator to separate the mark and space frequencies. The RF amplifier, mixer, and IF amplifier stages are similar to those of a normal communications superhet, although the bandwidths of these stages are reduced. The output from the last IF amplifier is fed to the input filter which accepts the mark and space frequencies but rejects unwanted interfering signals. The limiter passes constant amplitude mark and space signals to selective filters which separate the mark and space frequencies and pass them to the detectors. The positive (mark) and negative (space) pulse outputs are then amplified in the wideband pulse amplifier, the output of which is well-shaped pulses of + 80 volts suitable for operating a teleprinter relay. An emitter-follower provides a correct match to a 600-ohm line without distorting the pulses. To improve the reliability of FSK systems, two separate aerials and receivers are sometimes used with a common crystal controlled local oscillator used for both receivers.

## RADAR RECEIVERS

### Introduction

15. Radar receivers are invariably of the superhet type but with certain special characteristics which differentiate them from the normal communications receivers. These desired characteristics are:

- a. **High Gain.** High gain is necessary so that the weakest echo may be detected and gains in the order of 150 - 200 dB are common.
- b. **Low Noise Figure.** The advantages of high gain will be negated if the receiver noise figure is also high since a reasonable signal-to-noise ratio at the input will be degraded to an unacceptable value at the output. Noise figures for microwave radar receivers should be no greater than about 6 dB.
- c. **Bandwidth.** The bandwidth of a radar receiver will inevitably be a compromise since, whereas it must be wide enough to encompass the frequency spectrum of the transmitted signal plus any Doppler shift which might occur, it should also be relatively narrow to minimize noise. In pulse radar, the optimum bandwidth is the reciprocal of the transmission pulselength, and a receiver with such a bandwidth is said to be 'matched' to the pulselength. For CW radar, the optimum bandwidth is the reciprocal of the length of the 'pulse' generated in the Doppler filter as the radar beam scans through the target.
- d. **Automatic Frequency Control.** Radar transmitters, especially those using self-excited power oscillators, tend to drift in frequency, and some means of automatic tuning must be incorporated in the receiver so that it follows the drift of the transmitter.

### CW Receivers

16. In a CW radar, echoes are received while the transmitter is operating and it is, therefore, necessary (except for very low power systems) to employ separate aerials for transmission and reception, and elaborate measures are sometimes needed to minimize direct leakage of transmitter power into the receiver. However, a controlled amount of leakage entering the receiver along with the echo signal supplies the reference necessary for the detection of the Doppler frequency shift on the received echoes. The amount of isolation required depends on the transmitter power and the accompanying transmitter noise as well as the ruggedness and sensitivity of the receiver.



17. The receiver of a simple CW radar is, in some respects, analogous to a superhet receiver. It is termed a homodyne receiver or superhet receiver with zero IF. The function of the local oscillator is carried out by the leakage signal from the transmitter. This type of receiver suffers from poor sensitivity which is overcome by using a receiver with a non-zero IF. The reference signal in this case is derived from a portion of the transmitted signal mixed with a locally generated signal of frequency equal to that of the receiver IF. Since the output of the mixer consists of two sidebands on either side of the carrier, a narrow band filter selects one of the sidebands as the reference signal. This receiver is, therefore, sometimes called a sideband superhet receiver. The received signal ( $f_o \pm f_d$ ) is mixed with the reference signal ( $f_o \pm f_{if}$ ) to produce a difference frequency of ( $f_{if} \pm f_d$ ). After IF amplification the Doppler signals are resolved in a Doppler filter bank. This consists of a number of narrow band filters which together cover the range ( $f_{if} - f_{d \max}$ ) to ( $f_{if} + f_{d \max}$ ), where  $f_{d \max}$  is the maximum Doppler shift expected. Each filter is provided with its own detector and an output from any one of these indicates a target. An electronic switch is used for rapid examination of each filter detector in turn. The number of targets that the radar can resolve at any one time is equal to the number of Doppler filters.

18. The receivers employed in FMCW radars are similar to those of the simple CW radar. In the homodyne type, the receiver consists of a crystal mixer followed by a low frequency amplifier and a frequency-measuring device. A reference signal is necessary to be able to extract the Doppler shift and the range, and this is usually obtained by direct connection from the transmitter rather than by transmitter leakage, so that its magnitude can be more easily controlled. In the sideband superhet type of FMCW receiver, the output from the mixer is an IF signal of frequency ( $f_{if} + f_b$ ), where  $f_b$  is composed of the range frequency  $f_r$  and the Doppler shift  $f_d$ . The IF signal is amplified and applied to a balanced detector along with the local oscillator signal  $f_{if}$ . The output of the detector contains the beat frequency (range frequency and the Doppler velocity frequency), which is amplified to a level where it can actuate the frequency measuring circuits. The output of the low frequency amplifier is divided into two channels: one feeds an average-frequency counter to determine range and the other feeds a switched frequency counter to determine the Doppler velocity.

### Pulse Doppler Receivers

19. In a simple pulse Doppler system, a pulsed coherent transmitter is coupled to a common aerial via a TR switch. As in the CW system, the Doppler signals are extracted at an IF and targets are resolved in velocity by means of IF Doppler filters. However, whereas in the CW system a target signal which appears in the Doppler filters consists of a single frequency lying above or below  $f_{if}$  by the Doppler shift,  $f_d$ , in the pulse Doppler system the echo is pulsed and is therefore composed of a number of discrete components of pure CW separated in frequency by the PRF. If the echo signal is from a moving target, all components in this spectrum are shifted by the Doppler frequency and in the IF stage the central component has a frequency ( $f_{if} \pm f_d$ ). In order to resolve the target velocity without ambiguity only this component must be allowed to appear in the Doppler filters, and because of this rejection of other than the central frequency component, much of the echo power is lost. The fraction of the total power contained in the central component depends on both pulse length and PRF, an increase in either increasing the fraction. This reduced effective power degrades the signal-to-noise ratio compared to an equivalent CW system.

20. The signal-to-noise ratio of the system described may be restored if the operation of the receiver can be confined to the short period in each pulse cycle when the required echo arrives. The effect of this is to reduce the effective noise power to an extent comparable to the reduction in effective echo power, the net result being that the signal-to-noise ratio is recovered to a value similar to that of an equivalent CW system. This is achieved by the addition of a gating circuit which causes the receiver

to be opened for a short period,  $\tau_g$ , after an interval,  $t$ , following the transmission of each pulse. The duration of the range gate,  $\tau_g$ , is equal to or slightly greater than the pulselength,  $\tau$ , and the interval  $t$  is controlled so as to cause the gate to coincide with the time of arrival of the selected echo.

21. In a single-gate system, the gate must be swept across the inter-pulse period in order to search for a target. This not only increases the search time, but also reduces the target information rate if the system is to be able to deal with multiple targets. An alternative solution is to employ a number of fixed range gates which together cover the entire inter-pulse period. As each gate requires a separate bank of Doppler filters, the cost in complexity can be high. For example, a pulse Doppler radar having a duty cycle of .1 can have up to 9 range gates, each of which might feed as many as 500 Doppler filters. However, by using Fast Fourier Transform techniques, this number of filters can be accommodated on a single integrated circuit card.

### **Logarithmic Receivers**

22. A requirement of a radar receiver is that it shall not easily saturate, so that for example, a weak target echo will still be visible when superimposed on a strong clutter signal. This requirement may be met by adopting the successive detection principle in which a detector is connected across the output of each IF stage, the outputs of the detectors being added together in a delay line in such a way that the delay introduced between the outputs of successive detectors is equal to the propagation delay through the intervening IF stage. Each stage, in addition to feeding the next stage, thus makes an independent contribution to the final output of the receiver. If the input to such a receiver is steadily increased, the final IF will be the first to saturate. After a further increase in input the penultimate stage will saturate, and so on. In this way, the resistance of the receiver to saturation effects is greatly extended. In such an amplifier there is an approximately logarithmic relation between input and output amplitudes, and a receiver employing an IF amplifier of this type is, therefore, known as a logarithmic amplifier.

## CHAPTER 18 - TRANSMISSION LINES

Contents	Page
Introduction .....	1
Types of Transmission Lines .....	1
Characteristic Impedance .....	3
Standing Waves .....	4
Input Impedance .....	5
Matching .....	6
Transmission Line Losses .....	6

### Table of Figures

14-18 Fig 1 Types of Transmission Line .....	2
14-18 Fig 2 Variation of Voltage and Current with Line Length .....	4
14-18 Fig 3 Effect of Termination on Standing Wave Ratio .....	5
14-18 Fig 4 Input Impedance along a Short-circuited Line .....	6

### Introduction

1. A transmission line is any system of conductors by means of which electrical energy can be transferred from one point to another with negligible loss. In radio and radar systems, the majority of transmission lines are used to transfer RF energy either from a transmitter to an aerial, or from an aerial to a receiver. The loss in each case must be as small as possible. In the first case, any loss of energy is wasteful since it involves a needless use of power at the transmitter, while in the second case, a loss of RF energy could mean a serious loss of signals at the receiver input. Thus, in any transmission line system, steps must be taken to ensure that energy losses incurred are negligible compared with the magnitude of the energy being transferred.

2. As well as having low losses, a transmission line should not 'pick up' stray external voltages which would degrade the signal/noise ratio. This is particularly important in a receiving system, where the strength of the signal is relatively low.

### Types of Transmission Lines

3. All transmission lines consist of a conducting medium and a dielectric medium (which may be air). There will be electric and magnetic fields in both media, and it can be shown that the energy is transmitted along the line mainly by the fields in the dielectric, the conductors merely acting as guides. The four examples of transmission lines shown in Fig 1 represent the main types of line used in radio systems. At frequencies higher than the metric band (which includes most radar systems) waveguides are used, and these will be dealt with in Volume 14, Chapter 20.

4. **Open Twin Wire Feeder.** The open twin wire feeder consists of two parallel wires spaced a small fraction of a wavelength apart (less than  $\frac{\lambda}{10}$ ). The magnetic fields around the two conductors tend to cancel, and losses due to radiation are kept to a minimum. It has the following properties:

a. **Advantages:**

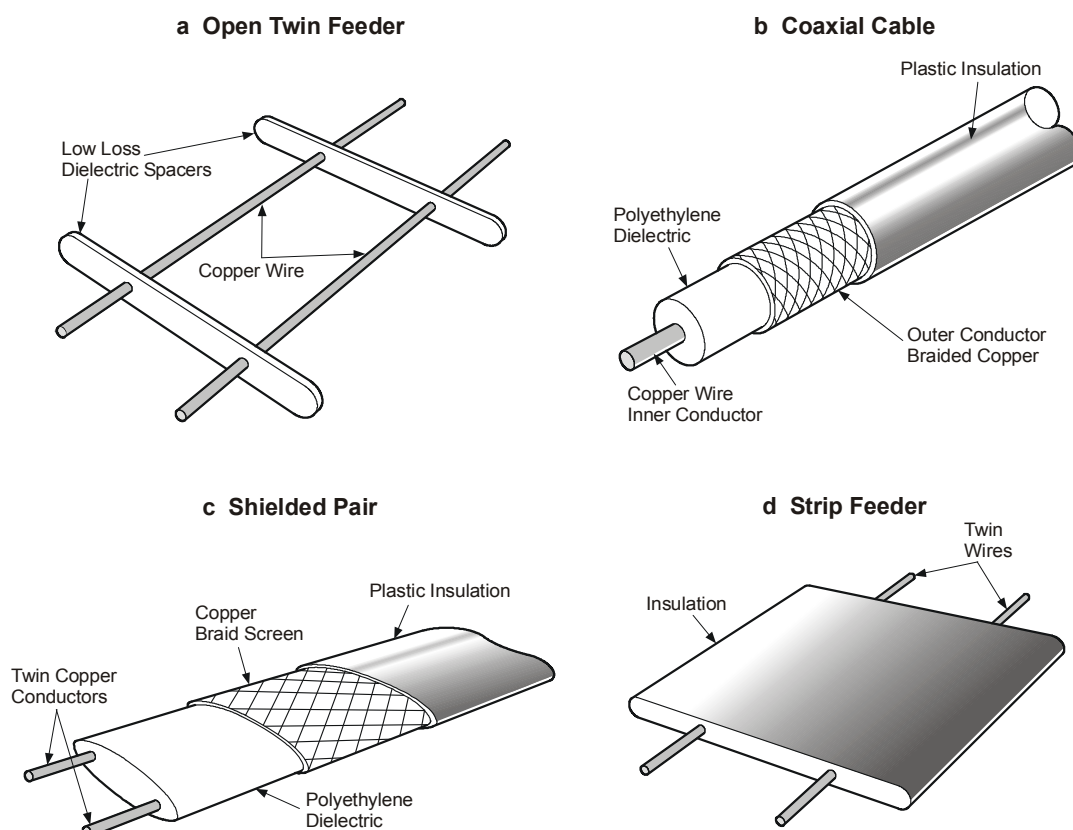
- (1) High transmitter power outputs can be handled without danger of breakdown of the air dielectric.
- (2) Standing waves (see paras 10 and 11) can be easily measured.
- (3) Maintenance of the line is relatively simple.
- (4) The line is balanced; that is, the impedance between wire and earth is the same for each wire. This is an important property when considering matching the line to the load.

b. **Limitations:**

- (1) It is bulky and rigid and can be used only on static installations.
- (2) At very high frequencies, the spacing becomes so small that, if high powers are being handled, there is a danger of 'flashover' between the wires. The upper frequency limit for high power installations is about 100 MHz.
- (3) It must be kept clear of the ground and walls.
- (4) Radiation losses limit the upper frequency to about 400 MHz. In practice, twin feeders cannot be used above 200 MHz.

c. **Applications.** It is the type of line invariably used for feeding balanced aerials for the HF band. It is usually run about 3 m above the ground and supported on insulators at intervals of 70 to 100 ft.

14-18 Fig 1 Types of Transmission Line



5. **Coaxial Cable.** Coaxial cable is a concentric type of twin wire transmission line. The inner conductor is held in the correct position relative to the outer conductor (braiding) by the use of insulating washers, spaced along the line at frequent intervals, or by completely filling the space between the conductors with a low-loss dielectric, eg polyethylene. Compared with the open wire feeder, its properties are:

a. **Advantages:**

- (1) The coaxial line is a screened cable. The fields are confined within the space between the inner conductor and braiding. Since the braiding may be earthed, negligible energy from the outside will penetrate into the cable circuit; there is also little radiation from the cable.
- (2) It is flexible and compact and can be buried in the ground.
- (3) It can be used at higher frequencies than open wire line since losses due to radiation are negligible.

b. **Limitations:**

- (1) Power handling capabilities are less than those of an open wire feeder.
- (2) It is an unbalanced line; this introduces additional problems when matching the line to the load.
- (3) Although it can be operated at higher frequencies than twin wire lines, it is subject to skin effect and dielectric losses which can cause a half power loss (3 dB) at 3,000 MHz, in three metres of line.

c. **Applications.** It is widely used for radio and radar purposes up to frequencies of the order of 3,000 MHz though waveguides tend to be used over about 1,000 MHz.

6. **Shielded Pair.** The shielded pair (or screened twin wire feeder) consists of two parallel conductors, mounted in a flexible braid screen, suitably insulated from each other and the screen by polyethylene. The metal braiding surrounding the polyethylene acts solely as a screen. In this way, the advantages of the coaxial feeder are combined with one advantage of the open wire feeder. The shielded pair is thus a screened and flexible line as well as being balanced. However, its power handling capacity is relatively small, and it has higher power losses for a given size and weight compared with coaxial cable. It is, therefore, used only for special applications where a balanced line is required.

7. **Strip Feeder.** The strip feeder consists of two conductors mounted along the edges of a strip of insulating material. It is extremely flexible but has a relatively small power carrying capacity. Therefore, it is usually used only for receiving systems. It also has applications in modern radar systems, in printed circuit form, in connection with microwave filters, receiver heads, printed aerial arrays and miniaturized microwave assemblies.

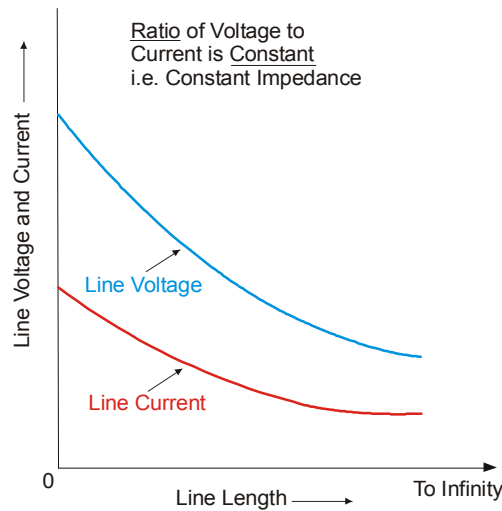
### Characteristic Impedance

8. The ratio of voltage to current, measured at intervals along a transmission line of infinite length, is found to be a constant, as shown in Fig 2. This ratio is known as the characteristic impedance ( $Z_0$ ) of the line. It is a complex quantity, but at very high frequencies the characteristic impedance is a pure resistance given by:

$$Z_0 = \sqrt{\frac{L}{C}} \text{ ohms}$$

where  $L$  = inductance per unit length, and  $C$  = capacitance per unit length.

**14-18 Fig 2 Variation of Voltage and Current with Line Length**



9. Instead of using inductance and capacitance per unit length to calculate  $Z_0$ , it is possible to use the physical dimensions of the line as follows:

a. **Twin Feeder.**

$$Z_0 = \frac{276}{\sqrt{k}} \log_{10} \frac{D}{r}$$

where  $D$  = distance between the centres of the wires,  $r$  = radius of wire,  
 $k$  = relative permittivity of dielectric (1 for air).

b. **Coaxial Cable.**

$$Z_0 = \frac{138}{\sqrt{k}} \log_{10} \frac{b}{a}$$

where  $a$  = radius of inner conductor,  $b$  = inner radius of outer conductor.

## Standing Waves

10. Although an infinitely long line is a physical impossibility, it is possible to get the characteristics of such a line to exist in a finite length if the practical limit is terminated in a resistance equal to the  $Z_0$ , of the line. If a transmission line is terminated in a resistance which is not equal to  $Z_0$ , the load cannot absorb all the energy being sent down the line and some of the energy will be reflected back to the generator. Although there are an infinite number of mismatch conditions, only the two extremes, open and short circuit, need be considered. Neither of these is capable of absorbing any energy, and so the incident wave is totally reflected. In order to reverse the direction of travel of an electromagnetic wave, only one component's direction should be reversed. In the case of a short circuit it is the electric field (voltage) component that is reversed, or phase shifted by  $180^\circ$ , whereas, for an open circuit, it is the magnetic field (current) component that is phase shifted by  $180^\circ$ . Thus, at a short circuit, incident and reflected electric fields cancel out to give a zero electric field, and incident and reflected magnetic fields reinforce to give a maximum magnetic field. The converse applies to the open circuit.

11. **The Standing Wave.** The resultant voltage distribution produced by the combination of the incident and reflected voltage waves is referred to as a standing wave. The positions of its maxima and minima on the line are fixed, and it rises and falls sinusoidally about these fixed positions. The

distance between an adjacent maximum (anti-node) and minimum (node) is a quarter wavelength. Voltage nodes coincide with current anti-nodes, and voltage anti-nodes with current nodes.

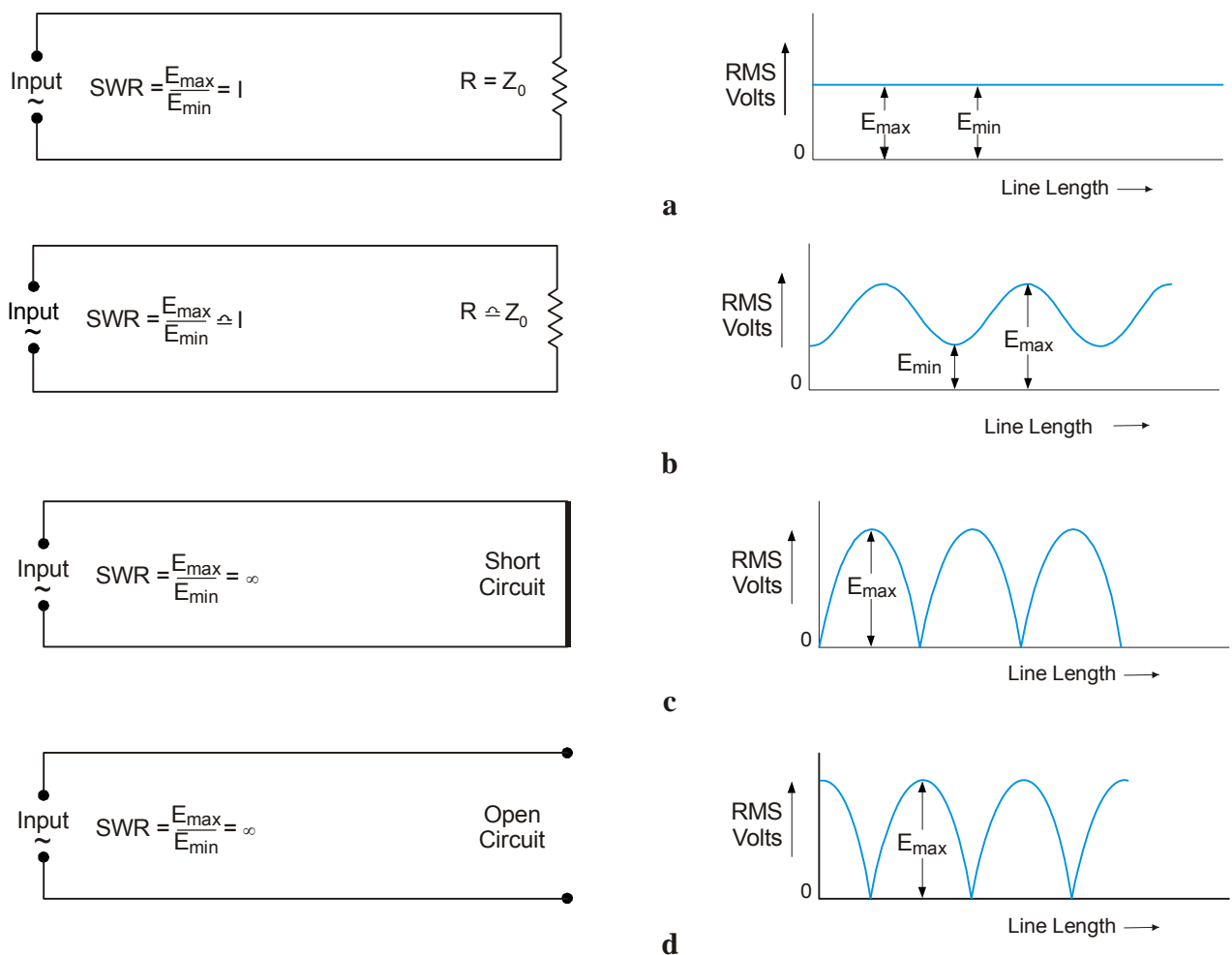
**12. The Standing Wave Ratio (SWR).** The range of variation between the maximum and minimum values of standing wave current or voltage gives an indication of the degree of mismatch, and is usually expressed in terms of the standing wave ratio as follows:

$$SWR = \frac{E_{\max}}{E_{\min}} \text{ or } \frac{I_{\max}}{I_{\min}}$$

where E max etc are rms values indicated by a meter.

**13. The Effect of Line Termination.** Fig 3 shows the effect of different line terminations on the standing wave ratio. When the line is correctly terminated in a resistive impedance equal to  $Z_0$ , there is no standing wave,  $E_{\max}$  equals  $E_{\min}$  and the SWR is unity (Fig 3a). With a slight mismatch ( $R$  slightly less or slightly more than  $Z_0$ ), the reflected wave is small,  $E_{\max}$  is only slightly greater than  $E_{\min}$  and the SWR is slightly greater than unity (Fig 3b). As the degree of mismatch increases, the SWR increases, and for the short-circuited (Fig 3c) or open-circuited (Fig 3d) transmission line, the SWR is  $E_{\max}/0 = \text{infinity}$ .

**14-18 Fig 3 Effect of Termination on Standing Wave Ratio**

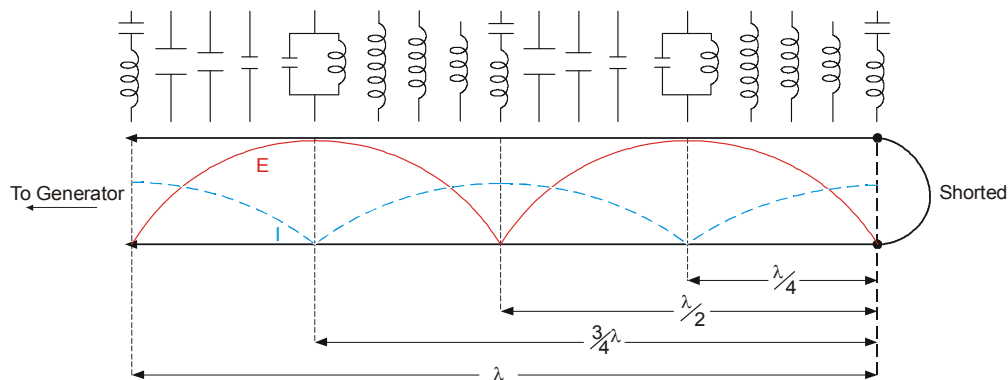


### Input Impedance

**14.** For a correctly terminated line, the input impedance is constant, no matter what the length of the line, and is equal to  $Z_0$ . With a mismatched line, however, the input impedance varies with length.

Fig 4 shows a short-circuited line, in which the impedance varies from minimum at the short circuit (corresponding to a series-tuned circuit at resonance), through a region of increasing inductive reactance to a maximum (corresponding to a parallel tuned circuit at resonance) a quarter wavelength from the short circuit, and then through a region of decreasing capacitive reactance, to a minimum again at the half-wavelength point. A short length of open or short-circuited line is, therefore, a most versatile component, and can act as inductor, capacitor, rejector circuit or acceptor circuit, depending on its length.

**14-18 Fig 4 Input Impedance along a Short-circuited Line**



### Matching

15. Where a load does not match the line directly, a short piece of suitable line can be used as an impedance matching device. It can be shown that a quarter wavelength of line impedance  $Z_T$  will match a load  $Z_L$  to the main transmission line of impedance  $Z_0$  when:

$$Z_T = \sqrt{Z_0 Z_L}$$

The impedance ( $Z_T$ ) required for this quarter wave transformer is achieved by either varying the diameter of its conductors, or, more commonly, by varying their distance apart.

16. **Stub Matching.** Where the load is reactive as well as resistive, the reactive component may be cancelled by a short-circuited line of equal but opposite reactance. This is termed 'stub matching' and may be used, either alone, or in conjunction with a quarter wave transformer.

### Transmission Line Losses

17. In their primary role as carriers of energy, transmission lines are subject to a number of losses. The most important of these are:

- a. Radiation loss.
- b. Skin effect loss.
- c. Dielectric loss.

18. **Radiation Loss.** Radiation loss affects twin feeders. The electromagnetic waves are merely guided by the line in the insulating medium surrounding it, and there is thus a tendency for the waves to escape into space. This is desirable in an aerial, but not in a transmission line. The fields of the wave may also induce currents in nearby conductors, providing another source of loss. The seriousness of radiation loss depends on the spacing between the waves in terms of the wavelength (the spacing should be less than  $\lambda/10$ ), and the use of twin feeder is restricted to frequencies below about 400 MHz, or lower if the line has to carry high power. Radiation loss does not affect coaxial line since the waves are contained within a closed reflecting surface and are thus unable to escape. Coaxial line may be



used up to about 3,000 MHz and is limited in its use by skin effect and dielectric losses. These two losses also affect twin feeders, but are overshadowed in their effect by radiation loss.

19. **Skin Effect.** Skin effect loss is basically a resistive loss (i.e. ohmic heating). The resistance of a conductor to alternating current increases with increasing frequency, since the current is caused to flow in a progressively decreasing depth of the conductor as a result of the magnetic fields set up by the currents. Thus, at frequencies of about 1 GHz, a thin-walled hollow tube has the same resistance as a solid tube of the same diameter. The effective increase of resistance is more marked for small diameter conductors than for large diameter ones.

20. **Dielectric Loss.** Dielectric loss arises partly from conduction currents, ie imperfect insulators, and partly from dielectric hysteresis. Dielectric hysteresis is the electrical equivalent of magnetic hysteresis and is a lag in electric flux behind the applied electric field. If the field is alternating, there is a heating effect in the insulator similar to the heating of a ferromagnetic material to which an alternating magnetic field is applied. Reduction of dielectric loss may be accomplished in one of two ways:

- a. **Improving the Dielectric.** The best dielectric is air, and air-spaced, or semi-air-spaced, lines are extensively employed. If large powers are to be handled, this may pose technical problems since the spacing between lines affects the radiation loss (see para 18).
- b. **Reducing the Volume of Dielectric.** Reducing the volume of dielectric reduces the power handling capacity of the line and also increases the skin effect losses.

## CHAPTER 19 - AERIALS

Contents	Page
Introduction .....	2
Properties of Electromagnetic Waves .....	2
Polarization .....	2
The Half-wave Dipole .....	3
Aerial Matching .....	4
Directional Properties of Aerials .....	4
Aerial Aperture .....	8
Image Aerials .....	8
VLF and LF Aerials .....	8
MF Aerials .....	10
HF Aerials .....	10
VHF and UHF Aerials .....	11
Aircraft Aerials .....	13
Broad Band Aerials .....	14

### Table of Figures

14-19 Fig 1 Electromagnetic Waves .....	2
14-19 Fig 2 Open Circuited Stub as a Series Resonant Circuit .....	3
14-19 Fig 3 Development of a Dipole .....	4
14-19 Fig 4 Radiation Patterns in Polar and Cartesian Coordinates .....	5
14-19 Fig 5 Field Strength Polar Diagrams of a Half-wave Dipole .....	5
14-19 Fig 6 Yagi Array with its Polar Diagram .....	6
14-19 Fig 7 Characteristics of Driven Arrays .....	6
14-19 Fig 8 The Rhombic Aerial .....	7
14-19 Fig 9 Marconi Quarter-wave Aerial .....	9
14-19 Fig 10 VLF and LF Aerials .....	9
14-19 Fig 11 Half-wave Vertical MF Aerial .....	10
14-19 Fig 12 Horizontal Centre-fed Half-wave Dipole for HF .....	10
14-19 Fig 13 Wide Band UHF Aerials .....	11
14-19 Fig 14 The Quarter -wave Ground Plane Aerial .....	12
14-19 Fig 15 The Turnstile Aerial .....	12
14-19 Fig 16 The Helical Aerial .....	13
14-19 Fig 17 Suppressed Aerials .....	13
14-19 Fig 18 VHF and UHF Aerials .....	14
14-19 Fig 19 Spiral Aerials .....	15
14-19 Fig 20 Log-periodic Aerial .....	15

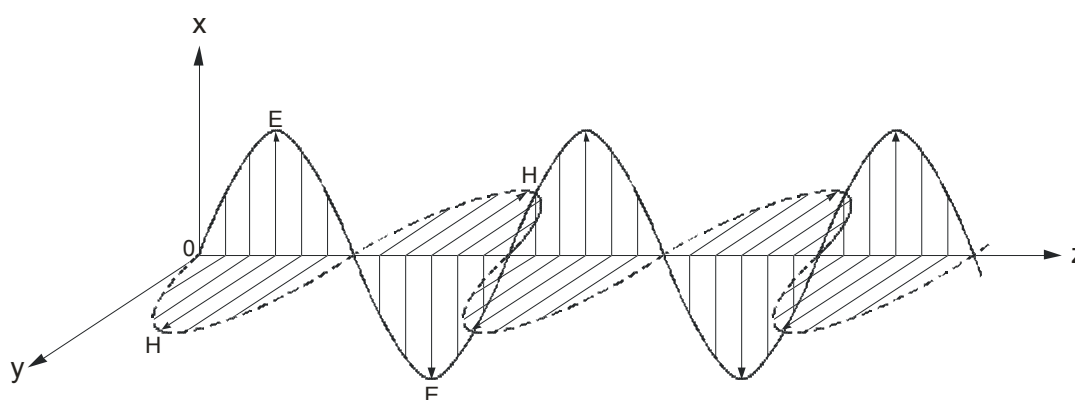
## Introduction

1. The purpose of an aerial is to act as a transducer between free-space propagation and guided-wave (transmission line) propagation. A transmitting aerial converts the electrical signals from a transmitter (wireless or radar) into an electromagnetic wave, which then radiates from it. A receiving aerial intercepts this wave and converts it back into electrical signals that can be amplified and decoded by a receiver.
2. In this chapter aerials will, in the main, be considered as radiating elements, however, the properties of a transmitting aerial apply equally well to a receiving aerial. This law of reciprocity means that many installations can use, in conjunction with suitable switching, a common aerial for transmission and reception.
3. Aerials used at centimetric wavelengths (microwaves) are dealt with as a separate subject in Volume 14, Chapter 20.

## Properties of Electromagnetic Waves

4. An electromagnetic (em) wave consists of mutually sustaining and moving electric (E) and magnetic (H) fields which are at right-angles to the direction of propagation, ie it is a transverse wave. The E and H fields are always at right-angles to each other in space as well as being at right-angles to the direction of propagation.
5. At any point in space the E and H fields are in phase with one another, and their variation as a function of distance in the direction of propagation is as shown in Fig 1. The velocity of propagation is 186,240 miles/sec, or  $3 \times 10^8$  metres/sec.
6. The ratio of E/H is a constant of 377 ohms, and is known as the impedance of free space.

14-19 Fig 1 Electromagnetic Waves



## Polarization

7. The polarization of an electromagnetic wave is defined as the orientation of the E field. This is convenient because the electric component is in the same plane as the linear radiating element of the aerial. Thus, a simple linear radiator orientated horizontally with respect to the earth will emit horizontally polarized waves. It is for this reason that aerials are often referred to as being horizontally or vertically polarized.

8. There are, however, certain types of aerial which emit a circularly or elliptically polarized wave. Under these conditions the directions of the E and H fields are not constant but rotate round the direction of propagation with constant amplitude, in the case of circularly polarized waves, and with varying amplitude in the case of elliptically polarized waves.

### The Half-wave Dipole

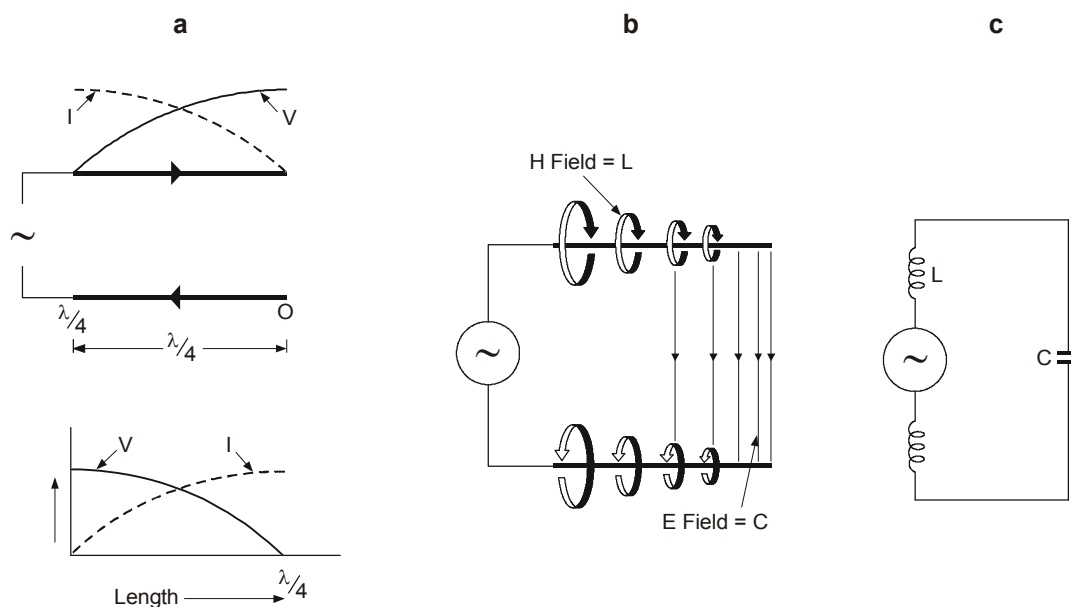
9. A simple transmitting aerial is a conductor, usually a wire, which is designed to radiate em waves as efficiently as possible. The simplest form of aerial is a half-wave dipole. This is a wire which is a half-wavelength long at the frequency of the current being carried by the wire, if the frequency is 30 MHz (10 metres wavelength), a half-wave dipole for use at this frequency will be 5 metres long.

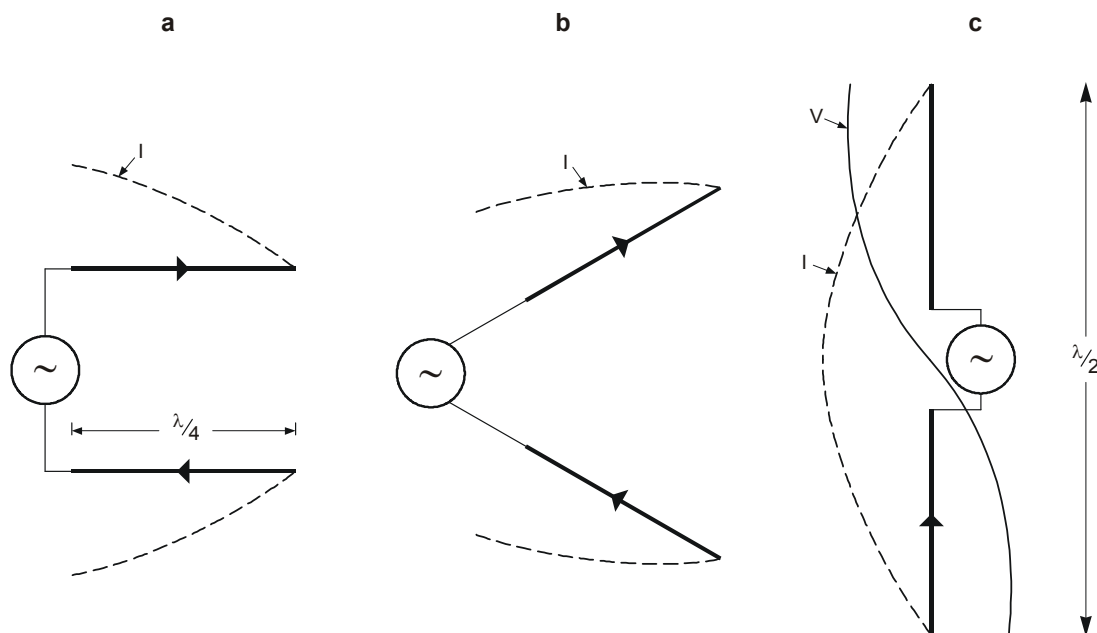
10. A half-wave dipole can be shown to work by comparing it with a quarter-wave section of open-circuited transmission line. Fig 2 shows that such a section may be considered as a series resonant circuit; it has a low input impedance and the current at resonance is maximum with the impedance at a minimum. The current and voltage distribution along the stub is shown in Fig 2a; note that the stub length is measured from the open-circuited end. The positions of maximum magnetic field (equivalent to inductance) and maximum electric field (equivalent to capacitance) are shown in Fig 2b.

11. The magnetic fields around the two conductors are caused by currents flowing in opposite directions. In the space between the conductors these fields reinforce each other, but away from the conductors the fields cancel; thus very little radiation occurs.

12. However, if the lines are opened out as shown in Figs 3b and c, the currents in the two conductors flow in the same direction and have a maximum value at the centre falling to a minimum value at the ends. There is now an open oscillatory circuit and em waves are radiated into space. This is the dipole aerial. The standing waves of voltage and current along the dipole are shown in Fig 3c. As the aerial is vertical the axes of the graph have moved through 90° compared with those of Fig 2a.

14-19 Fig 2 Open Circuited Stub as a Series Resonant Circuit



**14-19 Fig 3 Development of a Dipole**

### Aerial Matching

13. In transferring energy to the surrounding space in the form of electromagnetic radiation an aerial converts energy from one form to another just as a resistor converts electrical energy into heat energy when a current flows through it. The radiation resistance of an aerial is defined as that resistance which, when connected in place of the aerial, would dissipate the same power. For a half-wave dipole the radiation resistance is approximately constant at 73 ohms. In order to achieve maximum transfer of power to the aerial, the characteristic impedance of the transmission line must be matched to the input impedance of the aerial (for an aerial at the resonant length the input impedance is equal to the radiation resistance) and a number of devices, such as the Delta match, are available for doing this. In the Delta match, two points on the dipole are found where the aerial impedance matches that of the twin feeder transmission line at the point of connection. As the voltage between these two points on the dipole is practically zero, the centre of the aerial can be made continuous. The twin feeder transmission line is opened out to fan the delta. Radiation occurs from the delta which upsets the radiation pattern and reduces aerial efficiency.

### Directional Properties of Aerials

14. Aerials do not in general radiate isotropically, ie radiate fields of equal intensities in all directions. In the great majority of cases it is desirable to concentrate power in certain directions and to minimize it in others. An important aspect of aerial design concerns the means of distributing (or receiving) power via the aerial in such a way that the performance of a particular communications system shall be optimized, eg long distance point-to-point communication by HF skywave.

15. The directional properties of aerials may be expressed either in terms of power gain or in terms of angular concentration. The two methods express different aspects of the same property, gain being of primary significance when considering the transmission and reception of power, and angular concentration or beamwidth being significant in applications where direction-defining properties are involved, eg radio direction finding and radar.

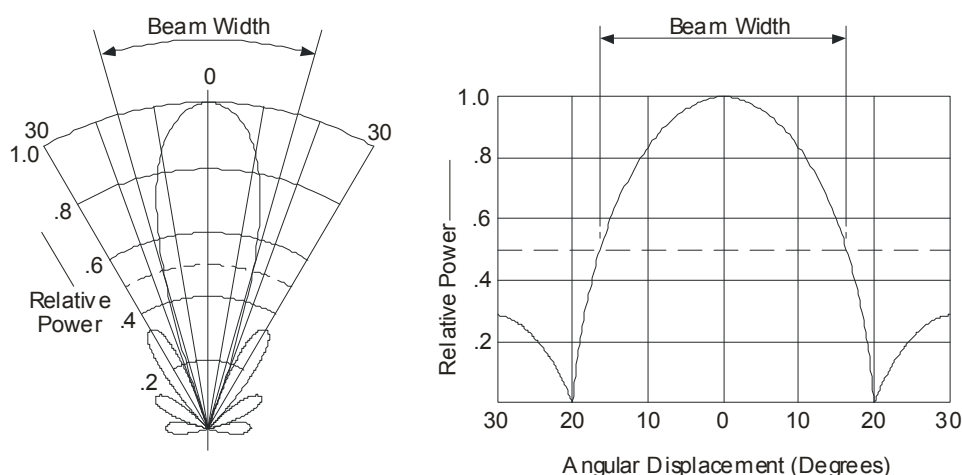
- a. **Gain.** The power gain of an aerial is normally defined as the ratio of the power it radiates in the direction of maximum concentration to that which would have been radiated from an isotropic source fed into equal power. (More generally, aerial gain may be referred to any defined

direction and hence may be greater or less than unity. By definition, the mean gain of any aerial is unity.) Aerial gain is usually expressed in decibels but in calculations involving the communications or radar equations the power ratio must be used.

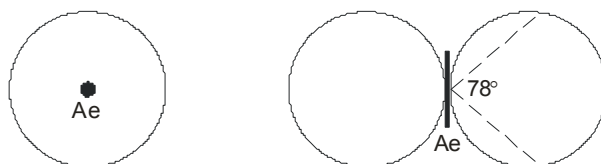
b. **Beamwidth.** The beamwidth of an aerial is defined as being the angle between the two directions in which the power radiated is one-half the power radiated in the direction of maximum concentration. The plane to which the measurement relates must be defined. Referred to field strength, beamwidth is the angle between the directions where the signal is  $0.707$ , ie  $1/\sqrt{2}$  of the maximum, since power is proportional to the square of field strength.

16. **Radiation Patterns.** To portray the directional characteristics of an aerial, graphs of relative power (or field strength) as a function of angle are plotted. Polar coordinates are normally used but for highly directional aerials it is difficult to interpret a narrow diagram and cartesian coordinates are used. The two types of radiation pattern graph for an aerial having a beamwidth of  $33^\circ$  are shown in Fig 4. To describe the directional properties fully it is necessary to portray the radiation pattern in two orthogonal planes. The radiation pattern of a dipole, shown in Fig 5, is omni-directional in a plane at right angles to its axis and has a beamwidth of  $78^\circ$  in any plane containing its axis. In most directional aerials the beam has an axis as opposed to a plane, and in radar aerials the beamwidths in orthogonal planes may sometimes be as small as  $0.5^\circ$ .

14-19 Fig 4 Radiation Patterns in Polar and Cartesian Coordinates



14-19 Fig 5 Field Strength Polar Diagrams of a Half-wave Dipole

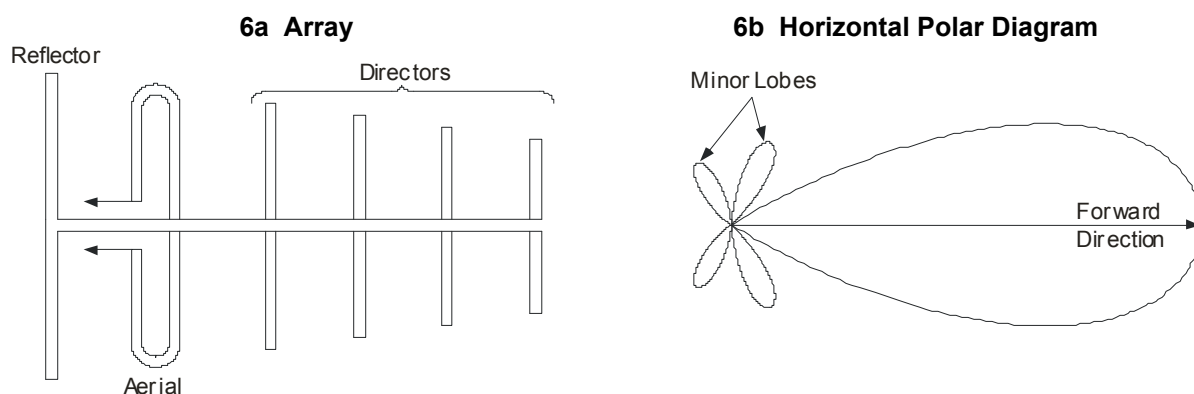


17. **Directional Arrays.** The dipole, which has an approximately omni-directional pattern, may be adapted in a number of ways to obtain directional properties. This may be achieved by the addition of parasitic elements, or by the combination of a number of driven dipoles.

a. **Arrays with Parasitic Elements.** Parasitic elements are conducting rods placed near a driven dipole. The transmitted energy of the dipole induces a current flow in the parasitic element, which then re-radiates the energy. By adjusting the lengths and spacings of the parasitic elements, either cancellation or reinforcement may be obtained in a given direction.

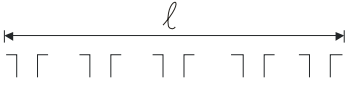

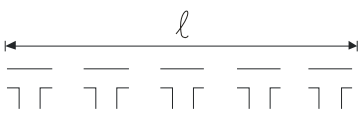

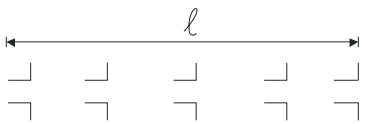

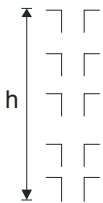

There are two types of parasitic element, the reflector and the director. The reflector's length is slightly greater than  $\frac{\lambda}{2}$  and when placed at the correct distance from the dipole the reflector cancels the radiation in a direction from the dipole to the reflector, but reinforces the radiation in the opposite direction. Directors are slightly shorter than  $\frac{\lambda}{2}$  in length and are placed, relative to the direction of maximum radiation, in front of the dipole. The most common aerial array having parasitic elements is the Yagi array which is formed from a driven dipole with at least one reflector and one director. The beamwidths in the two orthogonal planes are comparable, and by the use of additional directors (normally up to a maximum of about five) beamwidths down to about  $30^\circ$  may be achieved in both planes. The spacing of the elements is theoretically about  $\frac{\lambda}{4}$  but in practice is between  $0.1\lambda$  and  $0.15\lambda$ . The presence of parasitic elements reduces the radiation resistance of the dipole, and since this is undesirable because of matching problems, the dipole is usually of the 'folded' type whose radiation resistance is higher than that of a single dipole, and so the overall radiation resistance is maintained at an acceptable level for convenient matching. A Yagi array with a folded dipole is shown in Fig 6a and its polar diagram in Fig 6b.

14-19 Fig 6 Yagi Array with its Polar Diagram



b. **Multiple Driven Arrays.** Directional arrays are often used at HF and above. They have two or more driven elements instead of parasitic elements. Some examples of multi-driven arrays and their characteristics are given in Fig 7. In the broadside array the spacing between the elements is half a wavelength with the result that the phases add in directions at right angles to the array and cancel in directions parallel to the array. The broadside array can be made to radiate in one direction only by placing parasitic reflectors a quarter of a wavelength behind the driven elements. Such arrays give direction only in the plane of the array. It is often desirable to have an array which possesses directional properties in both the horizontal and vertical planes. This can be achieved by stacking a number of broadside arrays one above the other. This was a common technique in early metric radars and is the basis of modern electronically-steered aerials. In the 'end fire' array, the elements are spaced a quarter of a wavelength apart and are fed in phase quadrature. This causes the phases to add in one direction only.

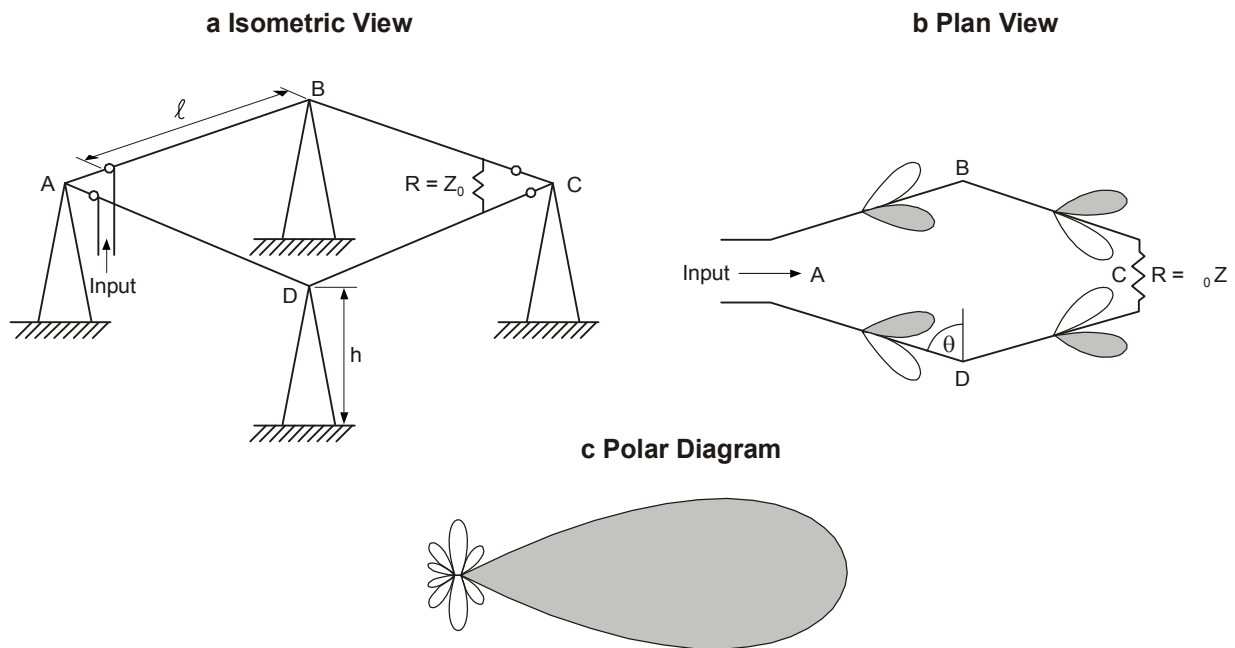
14-19 Fig 7 Characteristics of Driven Arrays

TYPE OF ARRAY	POLAR DIAGRAM (PD)	CHARACTERISTICS
 <p><b>a Linear Broadside Array (Plan View)</b></p>	 <p>Horizontal PD</p>	Aerials fed in Phase and spaced $\frac{\lambda}{2}$
 <p><b>b Linear Broadside Array with Reflectors (Plan View)</b></p>	 <p>Horizontal PD</p>	Aerials fed in Phase. Reflectors spaced $\frac{\lambda}{4}$ behind Aerials
 <p><b>c End Fire Array (Plan View)</b></p>	 <p>Horizontal PD</p>	Aerials fed $90^\circ$ Phase Lagging and spaced $\frac{\lambda}{4}$
 <p><b>d Stacked Array (Side View)</b></p>	 <p>Vertical PD</p>	Aerials fed in Phase and spaced $\frac{\lambda}{2}$ Vertically

**18. Travelling Wave Aerials.** The aerials so far considered in these notes have been standing wave (or resonant) aerials, on which the combination of incident and reflected waves forms standing waves. If the wire forming the aerial is terminated in a resistor equal in value to the characteristic impedance of the wire, then there will be no reflected energy and the only waves on the wire will be the incident waves moving towards the termination. These waves are called travelling waves and give this family of aerials its name. Travelling wave aerials are simple to construct and service, they possess good directional properties and, being non-resonant, can be used over a wide band of frequencies without being retuned. This is of considerable advantage when the system employs the sky wave and is required to operate on several widely spaced frequencies during day and night. The most common travelling wave aerial is the 'rhombic' which is widely used by ground stations for transmitting and receiving at HF. It consists of four wires forming a diamond or rhombus shape in the horizontal plane, as shown in Fig 8. It has a very wide bandwidth and a high gain. The direction of maximum gain is along the major axis of the rhombus towards the terminating resistor. The exact shape of both the vertical and the horizontal polar diagrams is determined by the leg length ( $\ell$ ), the tilt angle ( $\theta$ ) and the height of the aerial above the ground ( $h$ ). For example, as leg length increases, the beamwidth narrows and the angle of elevation decreases.

14-19 Fig 8 The Rhombic Aerial





### Aerial Aperture

19. A useful aerial parameter, which is related to gain, is the aerial receiving aperture. Aperture may be regarded as the effective area presented by a receiving aerial to the incident wave. The basic relationship between gain (G) and aperture (A) can be expressed as:

$$\frac{G}{A} = \text{Constant} = \frac{4\pi}{\lambda^2}$$

For a half-wave dipole, which has a gain of 1.635, the effective receiving aperture is approximately equal to  $\frac{\lambda^2}{8}$ .

### Image Aerials

20. In the basic description of aerials it is assumed that the aerials have always been placed so far above the earth's surface that any reflected energy from the surface (due to the original radiation from the aerial) is negligible. In practice aerials are seldom so far above the earth's surface, or the skin of an aircraft, and the actual directional characteristics exhibited are due to the vector summation of the direct and the reflected waves.

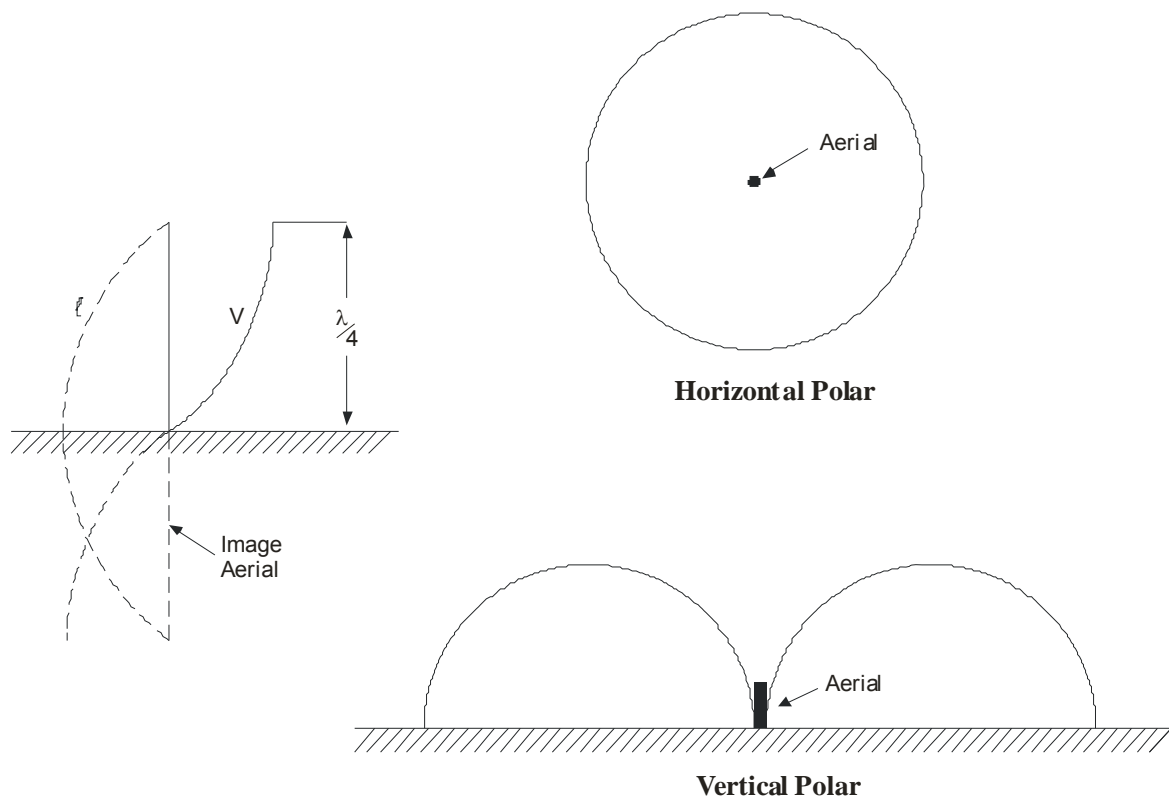
21. For the purpose of calculation, it is convenient to consider that the reflected wave is generated, not by reflection, but by an 'image' aerial located below the surface of the earth. The image aerial is so chosen that the joint action of the actual aerial and its image produces the same conditions in the space above the earth as exists with the actual aerial in the presence of the earth.

22. If the earth is considered to have a reflection coefficient of approximately unity, together with a  $180^\circ$  phase change, then the currents in the corresponding parts of the actual and image aerials are of the same magnitude.

### VLF and LF Aerials

23. The major problem encountered in the design of aerials for operation in the VLF and LF bands is the physical size of the elements involved. For example, at 100 kHz the wavelength is 3 km (nearly 10,000 ft) and so a half-wave dipole is out of the question. The Marconi aerial consists of a single conductor a quarter-wavelength long. It is mounted vertically and the transmission line is connected between its lower end and the Earth. The Earth, acting as a reflecting surface, creates an image of the aerial and the whole system behaves as if it were a half-wave dipole. The conductivity of the Earth surrounding a Marconi aerial must be high and if, as in the case of dry sandy soil, this is not so, then a metal mesh round the base of the aerial, called a counter-poise, is necessary. The polar diagrams for this type of aerial are shown in Fig 9. It should be noted that the radiation resistance is about half that of a half-wave dipole, i.e. about 36 ohms.

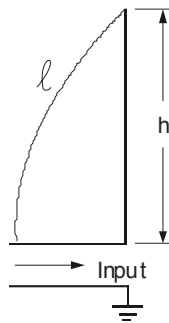
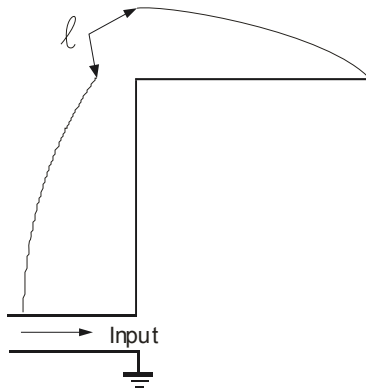
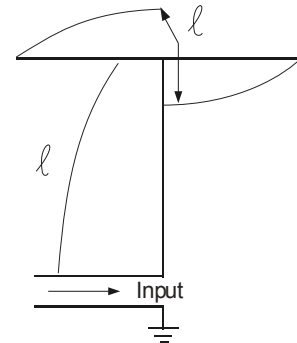
**14-19 Fig 9 Marconi Quarter-wave Aerial**



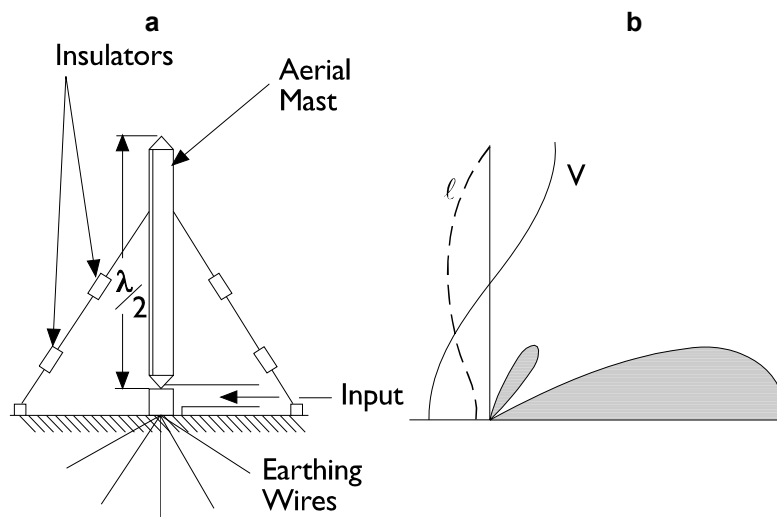
24. The Marconi aerial is not the complete answer however at VLF and LF since structural problems restrict heights in general to about 300 ft. Masts for aerials operating in these bands can therefore be only a fraction of a wavelength high and the efficiency of the aerials is correspondingly low. Furthermore, directional arrays would require massive structures and so omni-directional systems are generally used at these frequencies.

25. One method of increasing the radiation without increasing the mast height is shown in Fig 10. The current distribution for an aerial of height  $h$  is shown at a. By connecting a horizontal wire to the top of the aerial, the current distribution in the vertical (and most important) portion is as shown at b and the 'effective' height is thereby increased. This L aerial is very common at VLF and LF. A variation, known as the T aerial, is shown at c.

**14-19 Fig 10 VLF and LF Aerials**

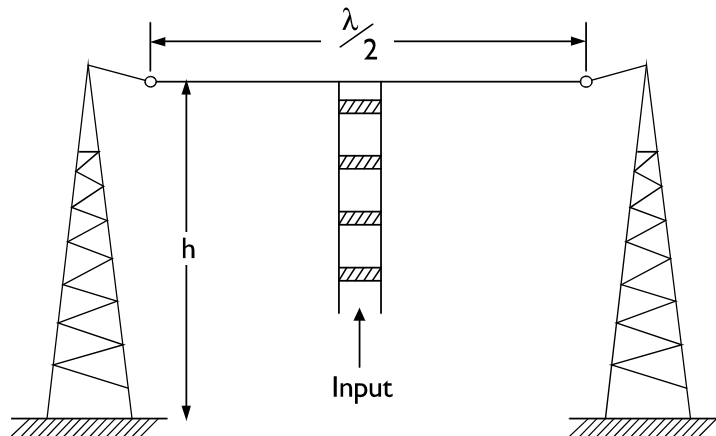
**a Vertical Aerial****b Inverted L Aerial****c T Aerial****MF Aerials**

26. At the lower end of the MF band the techniques employed at VLF and LF have to be used. Towards the upper end of the band, aerials half a wavelength long can be used. Fig 11 shows a typical half-wave vertical MF aerial. It is insulated at its base and energized with respect to an earth conductor system. The current at the base is therefore small and ground losses are low. The radiation resistance is high compared with the total loss resistance and so a high efficiency is obtained. The physical height of the aerial can be reduced by 5 to 10%, without affecting the radiation pattern to any marked degree, by the addition of a capacitance 'hat' on the top of the aerial.

**14-19 Fig 11 Half-wave Vertical MF Aerial****HF Aerials**

27. Communication in the band of frequencies between 3 and 30 MHz generally employs sky wave propagation. At the lower end of the band the simplest form of aerial used by ground stations is the horizontal half-wave dipole shown in Fig 12. It is rigged between two support masts and uses the Earth as a reflector. Maximum gain is obtained at high angles of elevation and no ground wave is present. This aerial is thus suitable for ionospheric communication using horizontally polarized waves.

**14-19 Fig 12 Horizontal Centre-fed Half-wave Dipole for HF**



28. In the HF band, increased ranges can be obtained by using highly directional arrays with the main beam directed at a shallow angle to the ionosphere. Rhombic and multi-drive arrays are used for this purpose. An example of the latter is the curtain array which consists of pairs of half-wave end-fed horizontal dipoles stacked vertically, the whole being suspended from two or more lattice masts. Reflectors are usually suspended behind the dipoles.

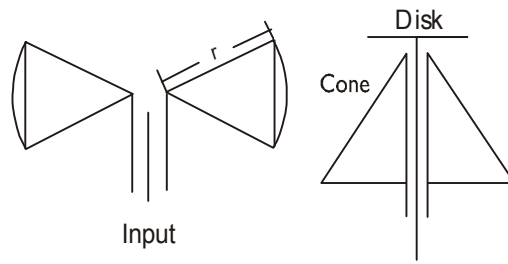
### VHF and UHF Aerials

29. Aerials for use in the VHF and UHF bands are generally made of hollow aluminium or copper tubing, and as the diameter of the elements can be increased, so the bandwidth becomes greater. Yagi arrays are in common use in these bands, particularly as receiving aerials, but since the wavelength is relatively small, aerials which are large in terms of wavelength and which often possess elaborate reflecting systems can be used. In these bands there is the problem of interference between the direct and reflected waves from the ground, but this can be minimized by placing the aerial on a mast at a height of several wavelengths from the ground.

30. Transmission in these bands is almost solely by means of the space wave and therefore the range depends on the heights of the transmitter aerials. Horizontal or vertical polarization can be employed. Some of the most common aerials at VHF and UHF are:

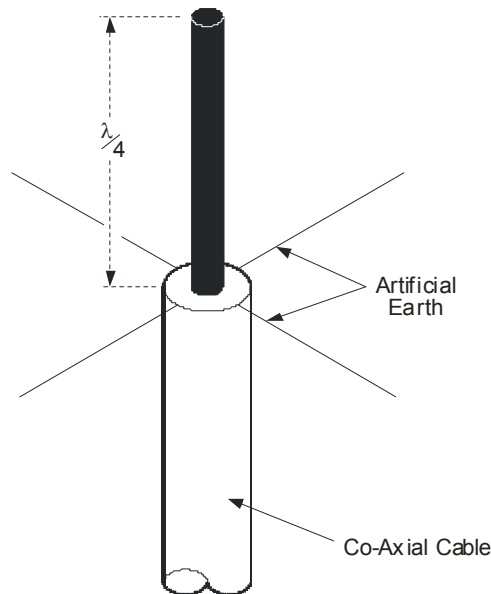
- a. **Biconical and Discone Aerials.** The biconical and discone aerials, shown in Fig 13, are widely used on UHF ground installations employed on air-to-ground communications. They will operate efficiently over the band for which the cone slant length,  $r$ , is between  $0.25\lambda$  and  $\lambda$ , thus giving frequency bandwidths of the order of 4:1 and greater. The cones are sometimes made in the form of a wire cage to reduce weight and wind resistance.
- b. **Quarter-wave Ground Plane Aerial.** As shown in Fig 14, this aerial consists of a  $0.25\lambda$  radiating element with four radial rods joined to the outer conductor of the coaxial feeder and forming an artificial earth. The aerial gives all-round radiation in the horizontal plane, and in the vertical plane the angle of elevation is determined by the length of the rods.

14-19 Fig 13 Wide Band UHF Aerials



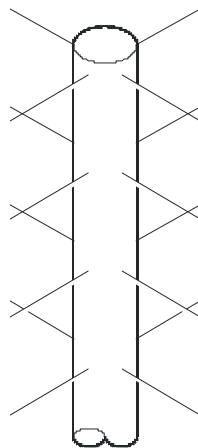
**a Bioconical Aerial      b Discone Aerial**

**14-19 Fig 14 The Quarter -wave Ground Plane Aerial**



c. **Turnstile Aerial.** The turnstile aerial, shown in Fig 15, gives all-round horizontally-polarized radiation in the horizontal plane suitable for broadcast transmission. The beamwidth in the vertical plane is reduced by stacking the crossed aerials in tiers. This increases horizontal radiation and decreases vertical radiation.

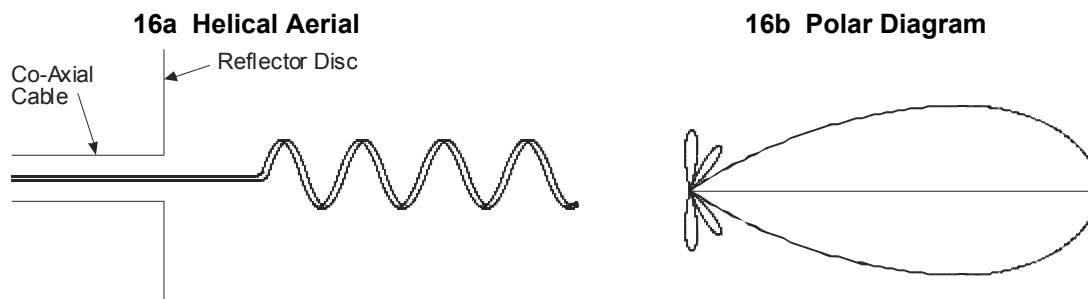
**14-19 Fig 15 The Turnstile Aerial**



d. **The Helical Aerial.** The helical aerial is of the travelling wave type. It provides a circularly-polarized beam of radiation which may be required in applications where the orientation of the

receiving aerial cannot be controlled, e.g. command guidance link to a guided missile. Its construction is shown in Fig 16. The diameter of the helix and the spacing between the turns must be carefully chosen to obtain the required polar diagram.

**14-19 Fig 16 The Helical Aerial**



### Aircraft Aerials

31. **LF, MF and HF.** Aerials employed on aircraft for use up to about 30 MHz are invariably restricted to dimensions which are shorter than one wavelength. In older aircraft carrying HF equipment this was partially overcome by a trailing wire of about 50 metres in length. An alternative is a simple wire aerial such as an inverted L or T rigged between the top of the tail fin and a small mast near the cockpit. When used for transmitting at relatively high altitudes, these wire aerials are prone to corona troubles due to the high voltage gradients combined with the low air pressure. When used for receiving they are liable to have a high noise level due to a phenomenon known as precipitation static. Higher aircraft speeds and the need to reduce precipitation static have necessitated the design of aerials known as 'suppressed aerials'. Suppressed aerials fall into two main classes:

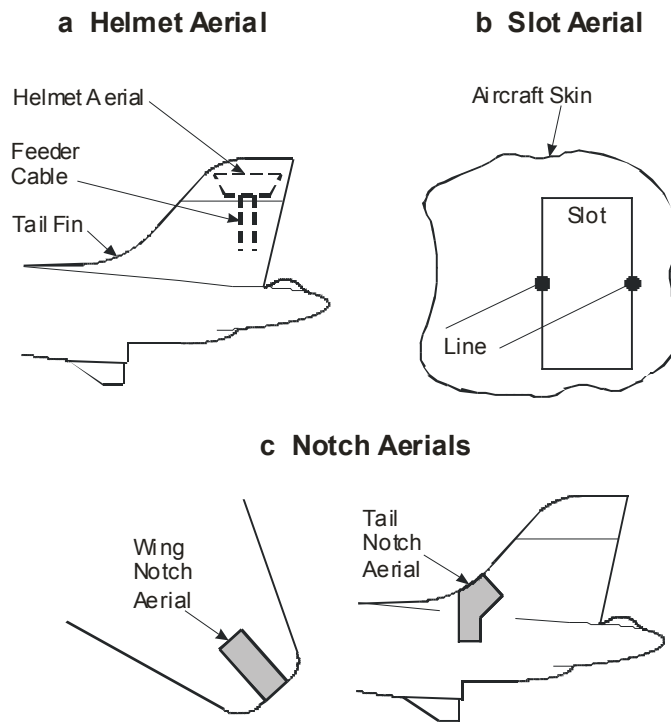
a. **Concealed Aerials.** In many cases these are conventional aerials but they are so constructed and mounted that they are concealed inside the aircraft. Typical of these are radar scanners mounted behind a radome nose section and the general purpose communication aerial of the 'helmet' type mounted on top of the tail fin and enclosed in a plastic cover. A helmet aerial is shown in Fig 17.

b. **Aerials Using the Aircraft Skin.** This type of suppressed aerial forms a resonant cavity within the aircraft and uses the metal skin of the aircraft as the radiating element. It may take one of two forms:

(1) **A Notch Aerial.** This is a notch cut out of the wing or tail and covered with insulating material to prevent the notch affecting the aerofoil. The RF power from the transmitter is fed to both sides of the notch and the RF voltage set up across the notch causes currents to flow in the outside skin of the aircraft and an electromagnetic wave is radiated. The notch aerial is shown in Fig 17.

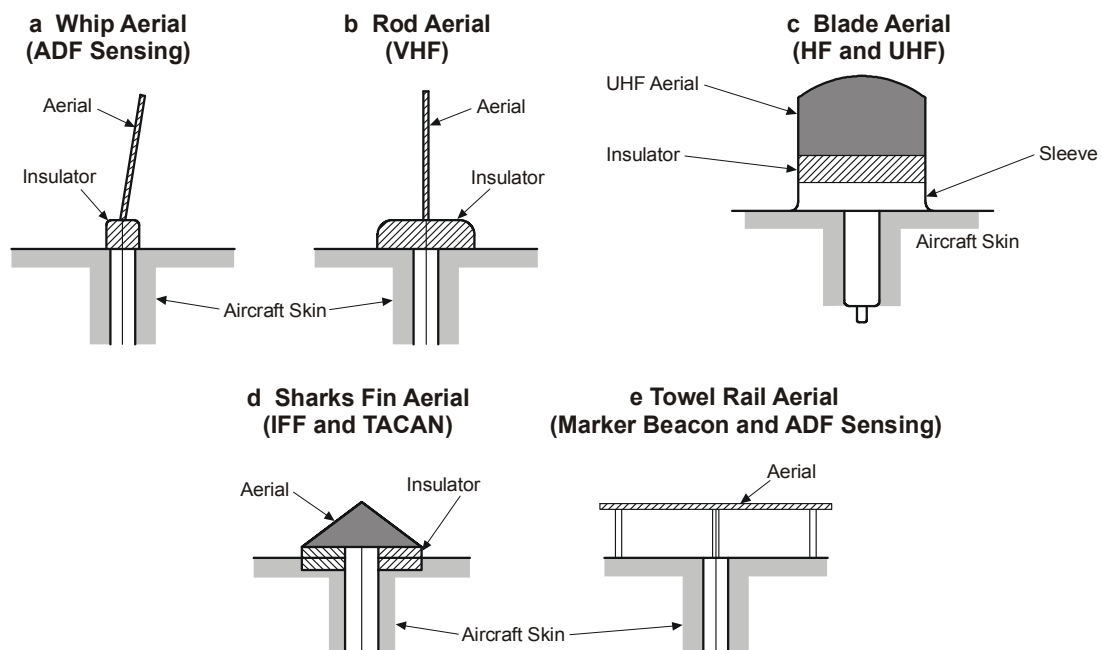
(2) **A Slot Aerial.** This is a half-wave length long slot cut in the aircraft skin. It differs from the notch in that the slot is resonant and functions in a manner similar to that of a half-wave dipole. The slot aerial is shown in Fig 17.

**14-19 Fig 17 Suppressed Aerials**



32. **VHF and UHF Aerials.** Aerials for aircraft use at VHF and UHF are generally easier to design than aerials for lower frequencies, simply because the wavelength is shorter. The most common types, with examples of their uses, are shown at Fig 18.

14-19 Fig 18 VHF and UHF Aerials



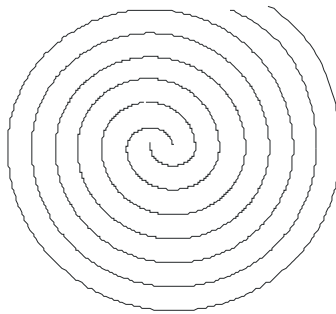
### Broad Band Aerials

33. In many applications it is necessary to have an aerial system whose performance remains virtually constant over a wide band of frequencies. Aerials based on conical shapes have a tendency to be broad banded. The discone aerial and the biconical aerial used at UHF (described in para 30) are examples of this. Unfortunately, the finite size of such aerials introduces end effects which limit the range of frequency insensitivity. Spiral aerials (Fig 19) are used extensively in passive radar

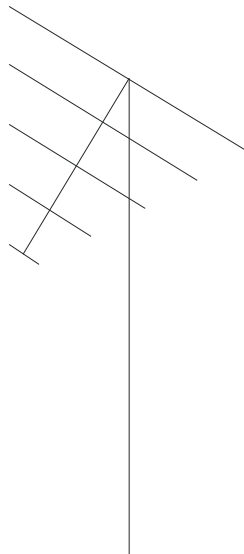
direction finding and semi-active missiles. By far the most successful broad band aerial is the log-periodic type (Fig 20). In the log-periodic aerial the distances between the active elements are designed to conform to a logarithmic expansion. Aerials of this type are in use for long distance HF communications and permit transmission with equal efficiency over the entire HF spectrum. They thus have bandwidths (ratio of highest and lowest frequency) of at least 10:1. Log-periodic aerials also find applications in ECM and radar.

34. The bandwidth of an aerial is also determined by its thickness. The thicker the aerial, the wider is the bandwidth.

**14-19 Fig 19 Spiral Aerials**



**14-19 Fig 20 Log-periodic Aerial**





## CHAPTER 20 - MICROWAVES

Contents	Page
Introduction .....	3
Resonant Cavities .....	3
Velocity Modulation .....	5
HIGH POWER AMPLIFIERS .....	5
The Klystron Amplifier .....	5
Multi-cavity Klystrons .....	6
The Travelling Wave Tube (TWT) .....	7
The Twystron .....	10
HIGH POWER OSCILLATORS .....	10
The Magnetron .....	10
The Carcinotron .....	12
LOW/MEDIUM POWER SOLID STATE DEVICES .....	12
Introduction .....	12
Point-contact Diode .....	12
Schottky-Barrier Diode .....	14
Tunnel Diode .....	14
P-I-N Diode .....	16
Varactor and Step Recovery Diodes .....	16
Metal Base Transistor .....	17
IMPATT Devices .....	17
TRAPATT (Trapped Plasma Avalanche Transit Time) Diode .....	18
Gunn Effect Oscillator .....	18
SURFACE ACOUSTIC WAVE (SAW) DEVICES .....	19
Introduction .....	19
Basic SAW Device Construction .....	20
LOW NOISE AMPLIFIERS .....	21
Introduction .....	21
The Parametric Amplifier .....	22
Masers .....	24
WAVEGUIDES .....	25
Introduction .....	25
Propagation in Waveguides .....	26
Waveguide Techniques and Components .....	27
MICROWAVE AERIALS .....	34
Introduction .....	34
Parabolic Reflectors and Feeds .....	34
Cassegrain Aerials .....	36
Lens Aerials .....	37
Flat Profile Aerials .....	39
Phased Array Aerials .....	39
Synthetic Aperture Systems .....	43

## Table of Figures

14-20 Fig 1 Development of a Resonant Cavity .....	4
14-20 Fig 2 Common Cavity Models .....	5
14-20 Fig 3 Klystron Amplifier – Schematic .....	5
14-20 Fig 4 Three-cavity Klystron Arrangement .....	6
14-20 Fig 5 Klystron Beam Focusing using Large Permanent Magnet .....	7
14-20 Fig 6 Travelling Wave Tube Arrangement .....	7
14-20 Fig 7 Travelling Wave Tube - Section of Helix Showing Electric Field Pattern .....	8
14-20 Fig 8 Travelling Wave Tube with Attenuator .....	9
14-20 Fig 9 Examples of Slow-wave Structures .....	9
14-20 Fig 10 Typical Magnetron Anode Block Structures .....	10
14-20 Fig 11 Cutaway Diagram of Magnetron .....	10
14-20 Fig 12 Electron Bunching in Magnetron .....	11
14-20 Fig 13 Point-contact Diode .....	13
14-20 Fig 14 Properties of a Point-contact Diode .....	13
14-20 Fig 15 Some Properties of a Tunnel Diode .....	14
14-20 Fig 16 Conventional Tunnel Diode Construction .....	15
14-20 Fig 17 Resonant-cavity Tunnel Diode Tuneable Oscillator .....	15
14-20 Fig 18 Variation of Capacitance in Varactor Diode .....	16
14-20 Fig 19 Basic Read Diode and Characteristic .....	18
14-20 Fig 20 Gunn Effect in Gallium Arsenide Semiconductor .....	19
14-20 Fig 21 Piezo-Electric Effect .....	19
14-20 Fig 22 Basic SAW Device .....	20
14-20 Fig 23 Tapped Delay Line .....	20
14-20 Fig 24 SAW Pulse Coding .....	20
14-20 Fig 25 FM SAW Device .....	21
14-20 Fig 26 A SAW Device for Pulse Compression .....	21
14-20 Fig 27 Principle of Parametric Amplifier .....	22
14-20 Fig 28 Degenerate Parametric Amplifier .....	23
14-20 Fig 29 Non-degenerate Parametric Amplifier .....	23
14-20 Fig 30 Three Level Cavity Maser .....	25
14-20 Fig 31 Travelling-wave Maser .....	25
14-20 Fig 32 Interaction of Two TEM Waves .....	26
14-20 Fig 33 Field Pattern Resulting from Two TEM Waves .....	26
14-20 Fig 34 Three-dimensional Illustration of $H_{10}$ Mode .....	27
14-20 Fig 35 Launching from a Coaxial Line by a Simple Probe .....	28
14-20 Fig 36 Coupling Slots .....	29
14-20 Fig 37 Crystal Detector Mounted on a Waveguide .....	29
14-20 Fig 38 Section of Choke Coupling .....	30
14-20 Fig 39 Various Types of Choke Flanges .....	30
14-20 Fig 40 Waveguide Rotating Joint .....	31
14-20 Fig 41 T-junctions .....	31
14-20 Fig 42 Branched Duplexer Principles .....	32
14-20 Fig 43 Balanced Duplexer .....	33
14-20 Fig 44 Parabolic Reflector .....	34
14-20 Fig 45 Some Typical Reflector Shapes and their Applications .....	35
14-20 Fig 46 Cosec <sup>2</sup> Dish .....	35
14-20 Fig 47 Feeds for Parabolic Aerials .....	35

14-20 Fig 48 Principle of the Cassegrain Aerial .....	36
14-20 Fig 49 Twist Cassegrain Aerial .....	37
14-20 Fig 50 Dielectric Lens Aerials .....	38
14-20 Fig 51 Metal Plate Lens .....	38
14-20 Fig 52 Spherical and Hemispherical Luneburg Lenses .....	39
14-20 Fig 53 Flat Profile Aerials .....	39
14-20 Fig 54 Phased Array Antennae .....	40
14-20 Fig 55 Beam-forming Network Using Additional Phase-shifters .....	42
14-20 Fig 56 Beam-forming Network using Tapped Delay Lines .....	42
14-20 Fig 57 Rotman Lens .....	43

## Introduction

1. Microwaves are a form of electromagnetic radiation with wavelengths approximately midway between those of light and radio waves, ie between 30 cm and 1 mm - with equivalent frequencies from 1 to 300 GHz. As with most divisions of the electromagnetic spectrum, these boundaries are somewhat arbitrary with no physical significance. Microwaves exhibit many of the same properties as both light and radio. The principle military applications of microwaves are in communications and radar. In radar systems, good azimuth resolution results from a narrow beamwidth. Short wavelengths mean that this can be achieved with relatively small aerials, which is especially desirable in airborne systems. In communications, the high frequencies allow high bandwidth signals to be transmitted. Microwaves are typically used for satellite communications links.

2. In order to utilize the microwave region of the spectrum, it is necessary to have devices which are capable of generating such frequencies and of amplifying them to useful power levels. Unfortunately, the conventional amplifiers and oscillators fail to perform satisfactorily at frequencies above 1 GHz, and it has been necessary to develop specialized devices which can operate at these frequencies. The properties of resonant cavities and the principle of velocity modulation are fundamental to an understanding of these devices.

## Resonant Cavities

3. At UHF, a short-circuited section of transmission line, called a 'lecher bar', may be used as the tuned circuit of an oscillator. Fig 1a shows the instantaneous electric and magnetic fields and the current in an oscillating lecher bar. The directions of the current and the fields change each half cycle. The relationship between the resonant frequency ( $f_0$ ), the inductance (L), and the capacitance (C) is given by:

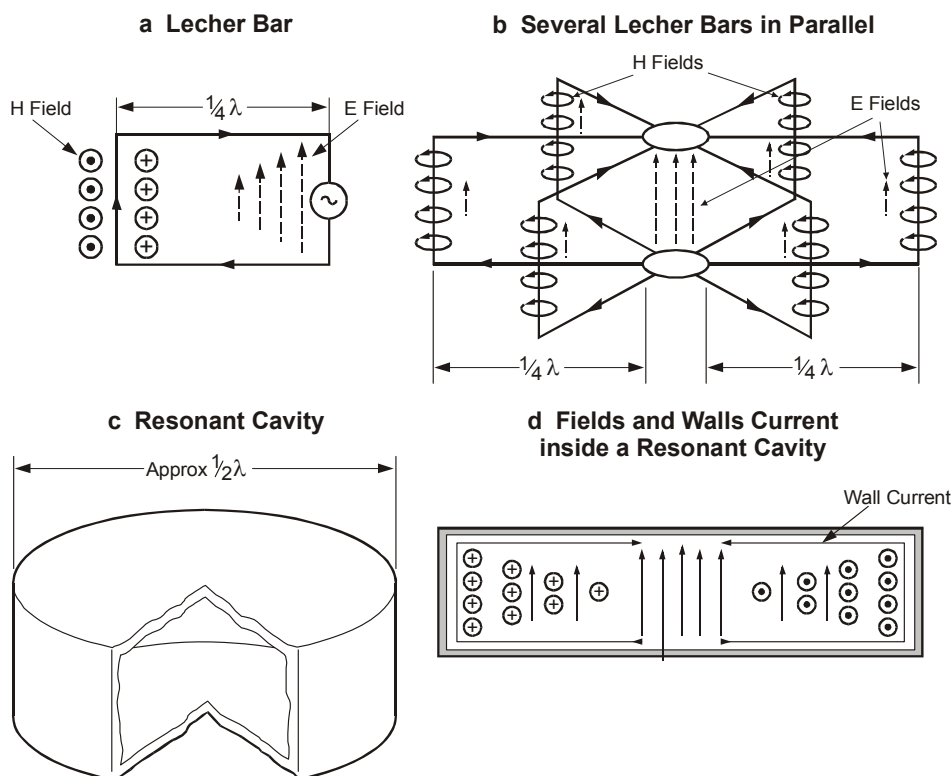
$$f_0 \propto \frac{1}{\sqrt{LC}}$$

If several lecher bars are connected in parallel (Fig 1b), the inductance is reduced by the same amount as the capacitance is increased; thus, the resonant frequency remains the same as for a single lecher bar. The total resistance is, however, decreased. If an infinite number of lecher bars are placed in parallel, a hollow cylinder or cavity is formed (Fig 1c). Little radiation can occur and, as the skin resistance is small, total losses are small.

4. The internal diameter of the cavity, which should be approximately half a wavelength, determines its resonant frequency. At centimetric wavelengths, the cavity dimensions are small enough for it to form an integral part of an amplifier or oscillator.

5. The instantaneous electric (E) and magnetic (H) lines of force present when a lecher bar oscillates alternate at a high frequency. The E lines may be associated with the circuit capacitance and the H lines with the circuit inductance. In a resonant cavity formed from an infinite number of lecher bars, the individual H lines round each lecher bar combine as in a solenoid and form closed loops, strongest around the circumference of the cavity and weakening to zero at the cavity centre. This H field always lies parallel to the cavity walls. The E lines combine to form a strong E field at the centre of the cavity, decreasing to zero at the cavity walls. During each half cycle the E field builds up to a maximum at the instant that the H field falls to zero. During the next half cycle the opposite happens, i.e. the E and H fields are  $90^\circ$  out of phase, as are the energies associated with the capacitance and inductance of a conventional tuned circuit.

**14-20 Fig 1 Development of a Resonant Cavity**

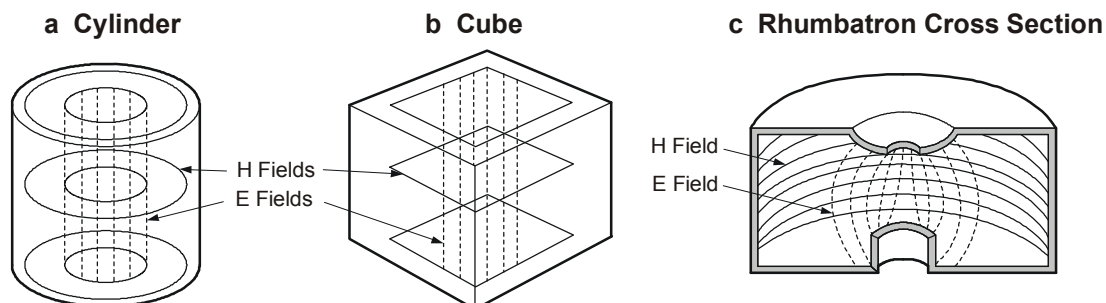


6. The fields shown in the cavity cross-section of Fig 1d are the fields at one instant of time. The electromagnetic energy in the cavity is oscillating at the cavity resonant frequency, the E and H fields changing direction every half cycle. The changing H field induces voltages in the cavity walls which produce wall currents. These wall currents are of the same frequency as, and always flow at right angles to, the H field. Due to skin effect, the currents flow only on the inner surfaces of the cavity and the outer surfaces can therefore be earthed without affecting the cavity operation.

7. A resonant cavity need not be cylindrical. Depending on the function of the cavity, it may be rectangular, spherical, or a modified cylindrical shape. In these cases, the field pattern (or 'mode'), which is formed inside the cavity when it is excited into oscillation, will differ from that shown in Fig 1d. Nevertheless, the E field will always be zero at the cavity walls, and the E and H fields will always be at right angles to each other and  $90^\circ$  out of phase.

8. Some common cavity modes are shown in Fig 2; the rhumbatron cavity (Fig 2c) forms the resonant cavity in the klystron. Electrons are passed through holes in the roof and floor of the cavity and, in order that the transit time of the electrons passing through the E field is short, the shape of the cavity is modified from a simple cylindrical shape, as shown.

**14-20 Fig 2 Common Cavity Models**



### Velocity Modulation

9. In a conventional amplifier, the device current is modulated in response to changes in the voltage existing in the input circuit; thus a sinusoidal input results in a device current that also varies sinusoidally. By making this current pass through a resistance in the external circuit, an output may be obtained which is identical in form to, but larger in amplitude than, the input. This amplitude modulation technique works well at VHF and below, but becomes progressively less effective as frequency increases, such that it is unusable above about 1 GHz. At these higher frequencies, a different technique called 'velocity modulation' is used.

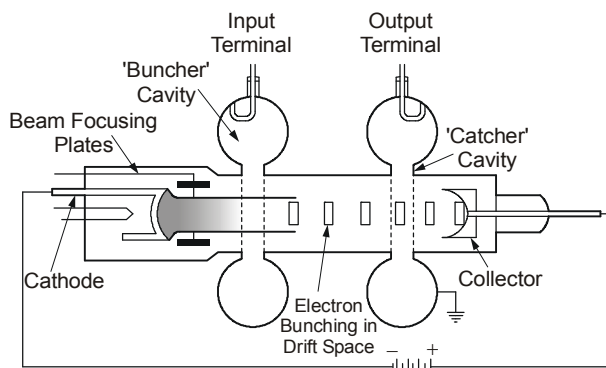
10. The principle behind velocity modulation is that of varying the velocity of the individual charge carriers that make up the device current, rather than their numbers. This results in bunches of charge being formed within the device. Using suitable circuitry, it is possible to use the bunch formation to produce an amplified output in a separate circuit. This technique is used in the klystron.

## HIGH POWER AMPLIFIERS

### The Klystron Amplifier

11. A schematic diagram of a klystron amplifier is shown in Fig 3. An evacuated tube contains an electron gun at one end. This sends a narrow beam of electrons to an anode collector at the other end. The beam passes through the narrow necks of two cavity resonators. The first cavity is known as the 'buncher', and it is at the same potential as the second cavity, the 'catcher'. Oscillations are induced in the buncher by inputting an RF signal by means of a short coaxial cable.

**14-20 Fig 3 Klystron Amplifier – Schematic**



12. The oscillating cavity generates an alternating electric field in the region through which the electron beam must pass; this field exerts a force on the electrons in the beam. Since the field is alternating, it accelerates electrons passing through the cavity during one half cycle, and decelerates electrons passing through during the other half cycle, i.e. it imposes velocity modulation. The result of the modulation is that the 'fast' electrons tend to catch up with the 'slow' electrons as they pass down the tube, with the result that electron bunches are formed at some distance from the buncher cavity (Fig 3). The second (catcher) cavity is placed so that the bunches are most pronounced just as they pass through it and, in doing so, they induce in it a strong oscillatory electric field, which has a polarity such as to slow down a bunch, thereby extracting energy from it. Half a cycle later, the direction of the electric field across the gap has reversed, and this field acts as an accelerator, but on the low density, non-bunched portion of the beam. Thus, many more electrons are slowed down than are accelerated, and there is a net exchange of energy from the beam to the catcher's output RF circuit, i.e. amplification.

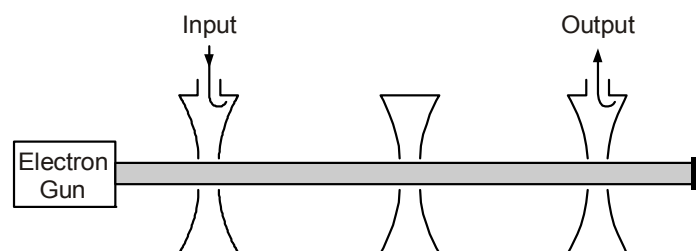
13. Klystrons of this type are readily available at I-band and above. The technique may also be used at frequencies below the microwave region. In the frequency range of 100 to 1,000 MHz, the size of the resonant cavity is such that quite large gap diameters can be used, together with similarly large diameter electron beams; this permits high powers to be achieved.

14. Typically, the two-cavity klystron described achieves power gains of 10 to 15 dBs, exhibits a rather low efficiency of about 40% and, in addition, has a narrow bandwidth (1 to 2%). In order to achieve better gains and efficiency, additional cavities are introduced.

### Multi-cavity Klystrons

15. It would, in theory, be possible to achieve higher gains by taking the output from one klystron and using it as the input to a second. This technique is used, in practice, by combining the two klystrons into one tube, as shown in Fig 4. The middle cavity is placed at the optimum distance from the buncher and acts as an additional velocity modulator. A gain of 40 dB is typical for a three-cavity device, but even higher gains can be achieved by introducing further cavities. Although, in principle, the number of cavities could be increased to any number, commonly either 4 or 6 cavities are used, giving gains of 50 and 110 dBs respectively, with efficiencies as high as 65%. The bandwidth may be widened, to a limited extent, by tuning each cavity to a slightly different frequency - a technique known as 'stagger tuning'.

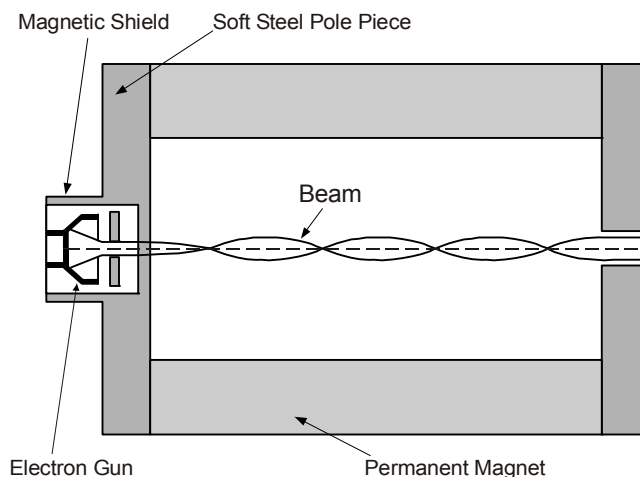
**14-20 Fig 4 Three-cavity Klystron Arrangement**



16. Multi-cavity klystron amplifiers are available at frequencies from about 400 MHz to 200 GHz. Peak powers produced range from a few kW to about 400 MW; average powers range from a few mW to about 175 kW. High power devices are, however, large and heavy and are not suitable for airborne applications. The kinetic energy of the electrons is converted into heat when they strike the collector, necessitating water cooling in some high power klystron amplifiers. This heat represents wasted DC energy and, to improve efficiency, the collector voltage may be decreased so as to reduce the electrons' kinetic energy.

17. **Beam Focusing.** In airborne applications, the weight of the device is important. The electron beam consists of negative charge, and repulsive forces will tend to spread the beam as it passes down the tube. The traditional way of preventing this is to provide a strong axial magnetic field by using a permanent magnet, as shown in Fig 5. The magnet is larger, and much heavier, than the rest of the klystron, typically 15 kg for a 1.5 kW device operating at 10 GHz. The use of samarium cobalt as the magnetic material is one way of reducing this weight by a factor of as much as 7 or 8; the material is, however, expensive. An alternative is to use a technique known as periodic permanent magnetic (PPM) focusing, which replaces the large magnet with a number of small ones.

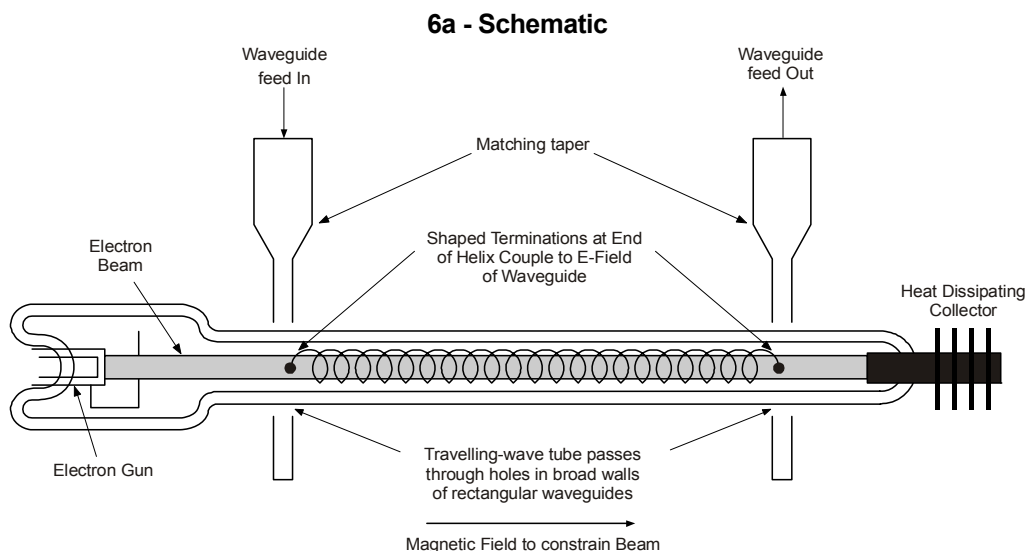
**14-20 Fig 5 Klystron Beam Focusing using Large Permanent Magnet**

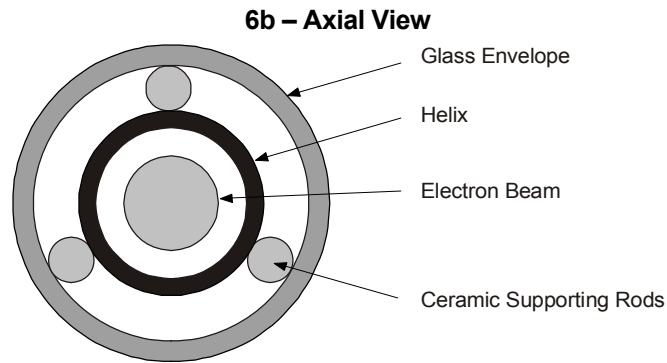


### The Travelling Wave Tube (TWT)

18. The main drawback of the klystron amplifier is its narrow bandwidth, resulting from the use of resonant cavities. The travelling wave tube (TWT) overcomes this problem. The principle of velocity modulation and charge bunching is still used, but the buncher and catcher cavities are replaced by a helix of wire; the arrangement is shown in Fig 6. Amplification is achieved by the exchange of energy between the electron beam and an electromagnetic wave travelling in the metal helix.

**14-20 Fig 6 Travelling Wave Tube Arrangement**

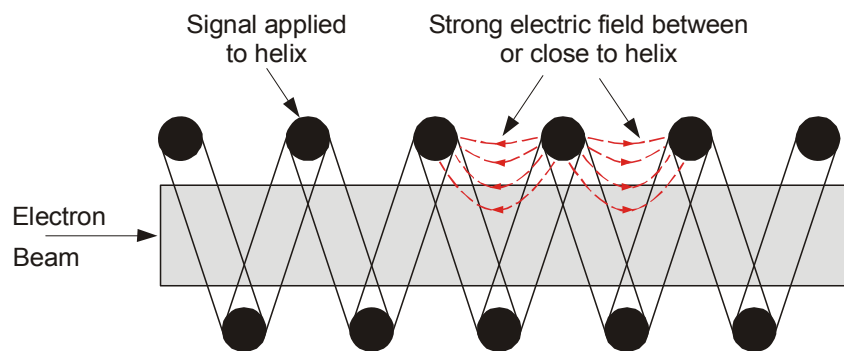




19. The electron beam passes down the centre of the tube and is kept in place by an axial magnetic field, generated either by a solenoid or by permanent magnets. The input signal is fed onto the helix at the electron gun end and it moves along the helical wire in the form of a voltage travelling wave; in principle, the same process as the sending of a signal down a transmission line. The signal travels around the wire of the helix at very nearly the speed of light. However, because of its greatly increased path length, its progress along the length of the tube is very much slower. The device is designed such that the travelling wave progresses along the tube at the same rate as the electrons in the electron beam and, since the travelling wave carries with it its own electric fields, the electrons in the beam and the travelling fields are able to interact with each other all the way along the tube.

20. Fig 7 shows a diagram of a portion of the helix and electron beam; examples of the alignments of two of the fields are illustrated. Any electron travelling down the tube will have alternating accelerating and decelerating forces applied to it. As with the klystron, velocity modulation and consequent charge bunching is achieved and, since the electrons travel with the field for the length of the tube, this bunching becomes more marked as the beam progresses.

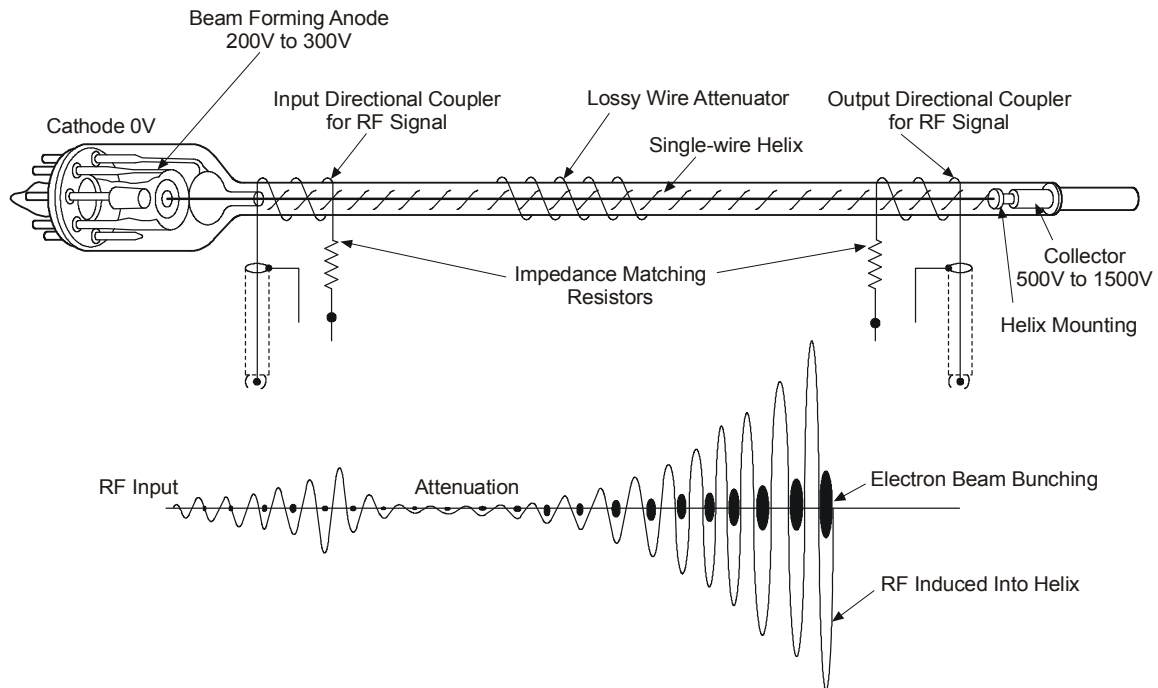
**14-20 Fig 7 Travelling Wave Tube - Section of Helix Showing Electric Field Pattern**



21. The electron bunches induce electric fields, and hence voltages, into the wire of the helix as they move down the tube. This voltage enhances the signal already present, ie amplification is achieved.

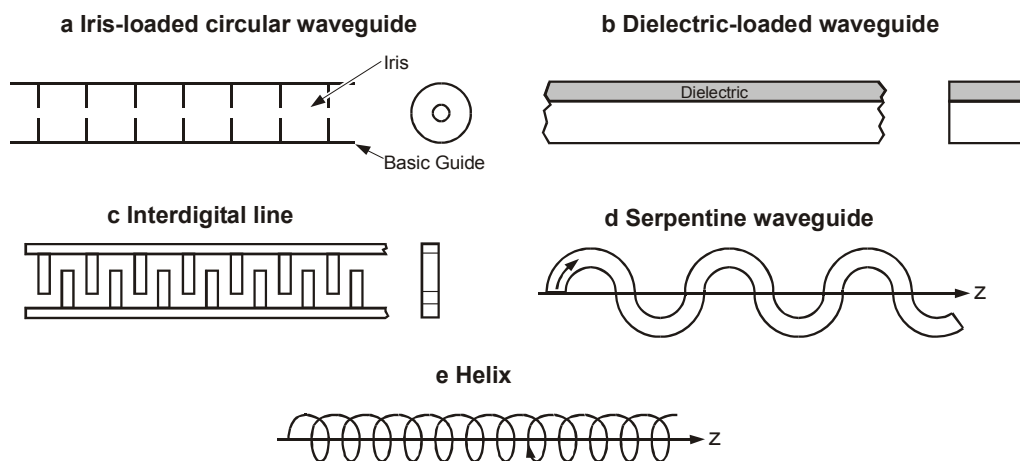
22. **Feedback.** Since the input and output are directly connected, it is possible for part of the output to be reflected back down the helix to the input, which can cause instability. It is, therefore, normal to place an attenuator near the wire to reduce any reflection. Although this also attenuates the proper signal, the electron bunches are not affected and proceed to re-induce a voltage in the wire to amplify the signal (Fig 8).



**14-20 Fig 8 Travelling Wave Tube with Attenuator**

23. **Slow-wave Structures.** The helix of wire in the TWT is designed to slow down the travelling wave to a speed along the tube commensurate with the electron beam. As such, it is an example of a slow-wave structure, and is the commonest structure used, but not the only one. Other forms of slow-wave structures are shown in Fig 9; all work on the principle of extending the path of the travelling wave.

24. **Frequency Range, Bandwidth, and Output Power.** TWTs exist at operating frequencies from 200 MHz to 50 GHz. Output power is very much lower than that of a multi-cavity klystron at, typically, up to a few kW; gains are in the order of 30 dB. The advantages of the TWT lie in its wide bandwidth (typically an octave), small size (length 30 to 60 cm), and low weight (a few pounds with PPM focusing). This makes it a particularly suitable device for airborne use, especially for frequency agile or jamming applications. It is also characterized by low noise and this makes it ideal as a communications amplifier.

**14-20 Fig 9 Examples of Slow-wave Structures**

## The Twystron

25. The multi-cavity klystron is characterized by high power, but narrow bandwidth, whereas the TWT gives wider bandwidth but with much lower power. Some applications, notably wide-band, high-power jamming, require the best features of both of these devices. The twystron aims to provide these characteristics by replacing the output cavity of a klystron with the output section of a TWT. The twystron is, therefore, a hybrid of the two devices and is able to provide peak powers in the MW range, bandwidths of 6 to 15%, and an efficiency of about 30%.

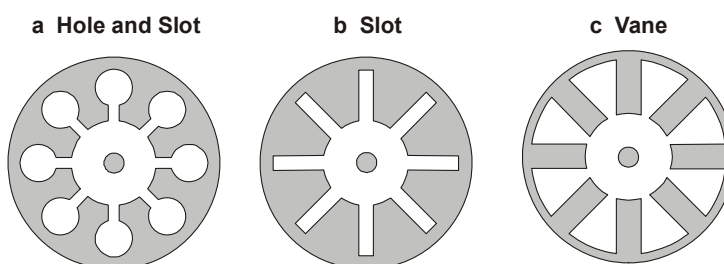
## HIGH POWER OSCILLATORS

### The Magnetron

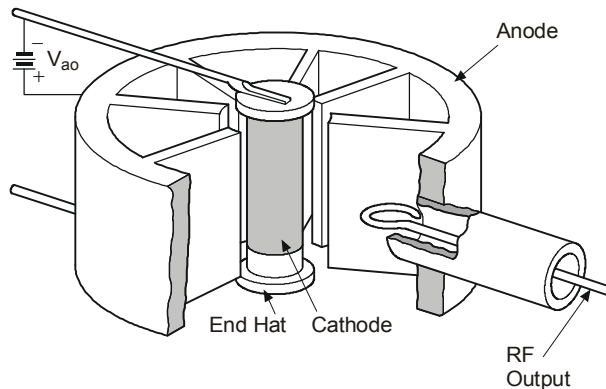
26. The high power oscillator in most radars is a device known as a 'magnetron'. It relies for its operation on the manner in which charged particles move in a magnetic field, experiencing a force at right angles to both the magnetic flux lines and their own motion. In a magnetron, electrons are emitted by a hot, cylindrical cathode, and are attracted by the high positive voltage of the anode, which surrounds the cathode in the form of a hollow cylinder. A magnetic field is established along the axis of the structure, which causes the electrons to move in a curved path towards the anode, the amount of curvature being dependent on the strength of the magnetic field. In practice, the field is arranged so that the electrons reach the anode after six or seven revolutions. It is the resulting motion of the electrons around the cathode that is used to produce oscillations

27. In a practical magnetron, the anode block is cylindrical in form, but is modified to include resonant cavities on its inner surface. Fig 10 shows some typical cavity shapes and Fig 11 shows a cutaway to illustrate the 3-dimensional structure.

**14-20 Fig 10 Typical Magnetron Anode Block Structures**

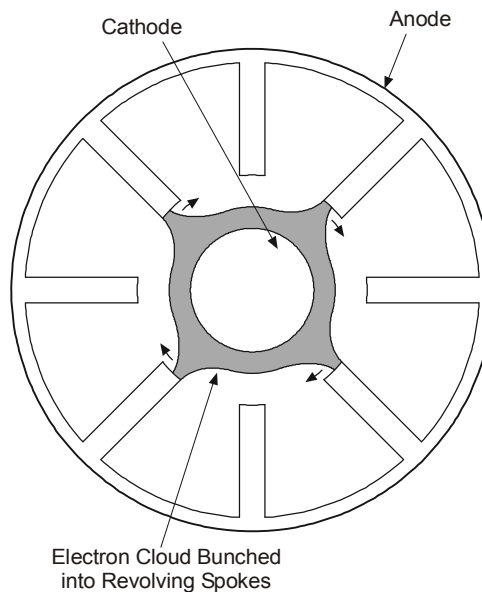


**14-20 Fig 11 Cutaway Diagram of Magnetron**



28. As the electrons spiral around the cathode, they pass the openings of the cavities and, if one of the cavities is oscillating at its resonant frequency, they will be velocity modulated by the field across its opening. Bunching will be initiated in the revolving electron cloud, and this in turn induces oscillations in the other cavities. The bunching will eventually reach the initially oscillating cavity; its oscillation will be enhanced and the bunching will become more pronounced. This interaction continues all the while that electrons are spiralling round in the device, the bunching coming to resemble spokes, as shown in Fig 12. In practice, there is always some oscillation present in all of the cavities as the result of thermal noise; this is sufficient to start the device. In order to use the power, it is tapped off from one of the cavities by means of a coaxial feed to a waveguide.

**14-20 Fig 12 Electron Bunching in Magnetron**



29. It is necessary to arrange the device so that the electron bunch reaches the original cavity at the right moment to enhance the oscillations present there. The normal arrangement, achieved by the choice of DC electric field and magnetic field strengths, is to have the bunch travel at such a rate that it moves from one cavity to the next in half a cycle. In this way, the bunch always experiences a retarding field, giving up its energy to the cavity.

30. The magnetic field is essential for the operation of the magnetron. If the field was removed, the electrons would travel directly outwards to the anode. In this situation, oscillation would cease and the current would rise, since the electrons would arrive at the anode in much greater numbers. With no oscillation, all the DC power would be converted to heat; this would be sufficient to melt the device. To obviate this danger, the magnetic field is always produced by a large and powerful permanent magnet, and this is the most bulky and weighty part of the structure.

31. The operating frequency of a magnetron is determined by the resonant frequency of the cavities, but a limited amount of mechanical tuning is possible by changing the cavity volumes with plungers. Inductive plungers in the cavity are used to raise the frequency, and capacitive plungers at the cavity entrances to decrease frequency.

32. Magnetrons are available, for both pulse and CW operation, with efficiencies between 30% and 60% for pulse operation, but lower for CW. In the pulse system, the magnetron is switched on and off by high voltage pulses (typically 20 kV) supplied by a separate circuit known as a 'modulator'. Output pulses are non-coherent and pulse lengths are normally in the range 0.1 to 5  $\mu\text{sec}$ .

33. The rated duty cycle (pulse length  $\times$  PRF) for pulse magnetrons is normally in the range of 0.005 to 0.00025, although, in some applications, a high duty cycle, approaching 0.5, is used. For low duty cycles, peak power outputs range from about 5 MW at UHF and D-band, to a maximum of about 200 kW at I-band. CW magnetrons generally have lower power outputs, typically 200 W at D-band to 1 W at G-band.

### The Carcinotron

34. The carcinotron is a travelling wave oscillator employing a slow wave structure. The field pattern of an electromagnetic wave travels at a certain velocity, known as the 'phase velocity', whereas the energy which it carries travels at a different velocity, known as the 'group velocity'. In the case of the TWT, these two velocities are in the same direction, but it is possible for them to be sensed in opposite directions. This is the case in the carcinotron, which is, accordingly, classified as a backward wave oscillator (BWO).

35. An electron beam is directed along the structure in the same direction as the phase velocity, and oscillations are possible at a frequency at which the phase velocity of the wave is synchronous with the velocity of the beam. The phase velocity of the electromagnetic wave changes with frequency, and so the frequency of oscillation depends on the velocity of the electron beam. Carcinotrons can therefore be tuned over a wide range of frequencies by varying the voltage of the electron beam. There are two types of carcinotron:

- a. **O-type.** The O-type of carcinotron is similar to the TWT, and the slow wave structure is usually a helix. It is normally a low power, low efficiency tube, principally used for receiver and test equipment applications.
- b. **M-type.** In an M-type carcinotron, the electron beam moves under the influence of orthogonal electric and magnetic fields. The device resembles a magnetron, in that it has a circular, rather than linear, geometry in order to reduce the size of the magnet. The efficiency of M-type carcinotrons is relatively high, and this, combined with wide range electronic tuning, makes them attractive for countermeasures or jamming applications.

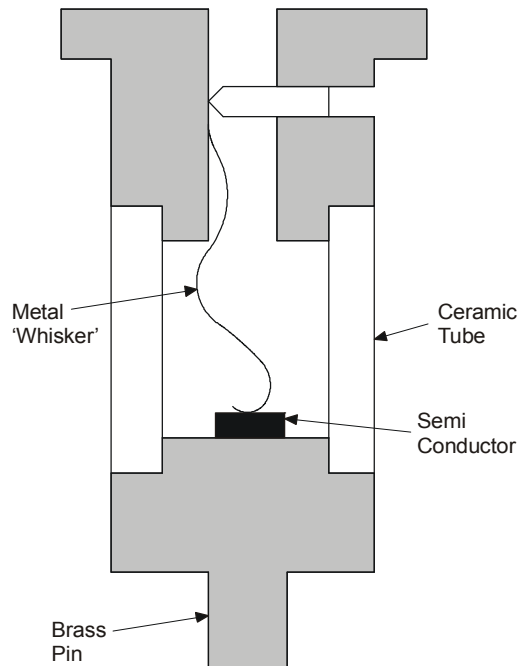
## LOW/MEDIUM POWER SOLID STATE DEVICES

### Introduction

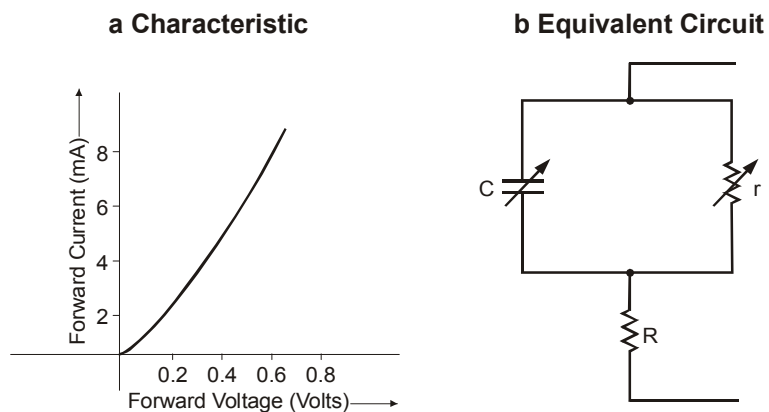
36. Whereas the high power requirements for the output stages of radar transmitters are catered for by the thermionic amplifiers and oscillators described above, radar receivers and microwave links in the low and medium power region normally use solid-state devices. These are far less noisy, have lower DC power requirements, and are also smaller, lighter, more reliable, and more efficient.

### Point-contact Diode

37. The point-contact diode, illustrated schematically in Fig 13, consists of a metal 'whisker' which presses on a small crystal of p- or n-type semiconductor. A diode made of p-type silicon with a tungsten whisker can handle very small signals, and is easily damaged by high power. In some cases, welded or gold bonded contacts are used. Crystals of gallium arsenide can withstand high temperatures and can operate at frequencies up to 100 GHz.

**14-20 Fig 13 Point-contact Diode**

38. Fig 14 shows the rectifying properties of a point-contact diode and its equivalent circuit. The resistance ( $r$ ) is the variable 'barrier' resistance of the contact and varies with the applied voltage, thus enabling the diode to be used as a mixer. The contact capacitance ( $C$ ) also varies with the applied voltage and is very low (approximately 0.2 pF) so that, at microwave frequencies, it does not short circuit  $r$ . The series resistance ( $R$ ) is the constant ohmic resistance of the semiconductor.

**14-20 Fig 14 Properties of a Point-contact Diode**

39. Point-contact diodes are fairly robust but, when used as microwave mixers, their life depends on the effectiveness of the TR device which protects them from the high transmitter power. The TR cell takes a short time to break down completely after the transmitter pulse has started, and a spike of energy leaks to the mixer diode during this period. If the power in the spike is too high, the metal to semiconductor contact may overheat and burn out. Over a period of time, even comparatively small leakage powers can cause the contact to change its rectifying properties, and the diode becomes inefficient and noisy. This can be detected by a deterioration in the overall performance of the radar, or by a decrease in the back-to-front resistance ratio of the diode.

40. **Backward Diode.** The backward diode is a form of point-contact diode in which the whisker is coated with semiconductor material and bonded to a crystal of n- or p-type germanium. This semiconductor-to-semiconductor contact operates better than the point-contact diode at low signal levels.

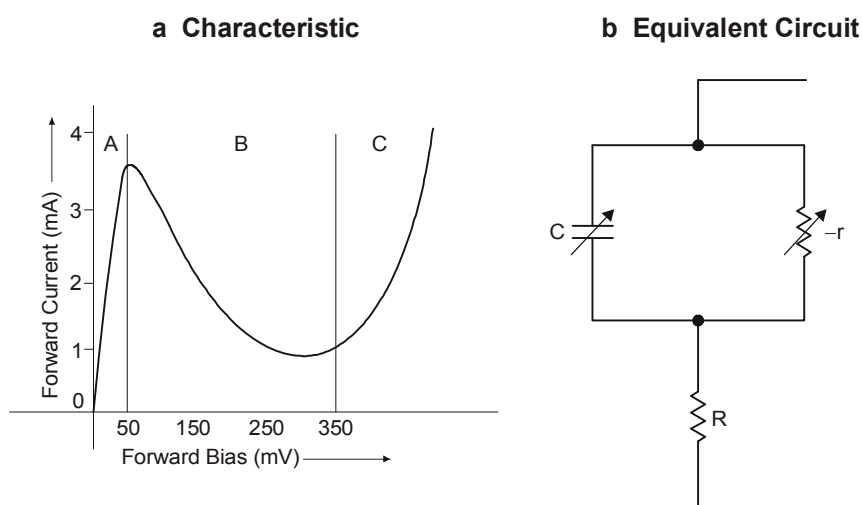
### Schottky-Barrier Diode

41. The Schottky-Barrier diode uses a metal-to-semiconductor junction, and consists of a thin metal film deposited on a layer of n- or p-type semiconductor. Because of the larger junction area, its capacitance is greater than that of the point-contact or backward diodes (about 1 pF), but it is not easily damaged by high power. Its characteristics are fast switching speed, low forward turn-on voltage, low stored charge, low reverse leakage currents, and high rectification efficiency; hence it is used in circuits for high and low level detection, mixing, UHF modulation, pulse-shaping, voltage damping, and pico-second switching (in computers).

### Tunnel Diode

42. The tunnel diode is a semiconductor junction made of heavily doped germanium. It has a very narrow depletion layer between the p- and n-type materials and current-carrying charges can 'tunnel' through the layer at low values of forward bias. From about 0 to about 50 mV, the forward current due to these 'tunnelling' carriers rises fairly sharply (region A in Fig 15a) but, as the voltage is increased above about 50 mV, the tunnelling effect decreases and the forward current falls (region B). At about 350 mV, normal semiconductor action takes over, and the current again increases with an increase in forward voltage (region C).

14-20 Fig 15 Some Properties of a Tunnel Diode



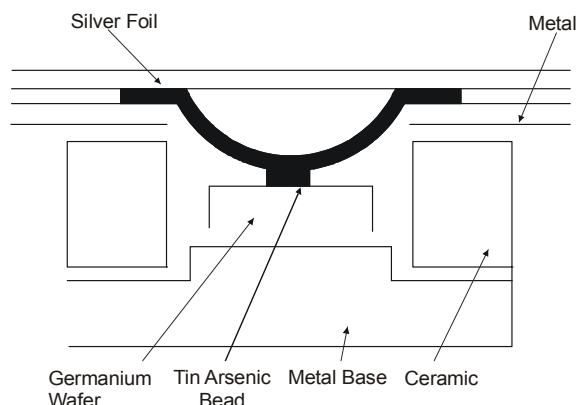
43. By operating the diode in the region where the voltage/current curve is negative, it can be used as a negative resistance oscillator or amplifier. When employed in an oscillator circuit, the positive resistance of the circuit is completely cancelled by the negative resistance of the diode. In an amplifier circuit, the positive resistance is partly cancelled, and the input applied to the circuit is amplified.

44. The equivalent circuit of a tunnel diode (Fig 15b), is similar to that of a normal semiconductor diode, but the variable barrier resistance becomes negative ( $-r$ ). The constant resistance ( $R$ ) is due to the semiconductor material and the capacitance ( $C$ ) depends on the junction area and width. Thus,  $C$  is fairly high (about 10 pF) and, at microwave frequencies, the tunnel diode presents a very low impedance.

45. The transit time of the carriers through the very narrow depletion layer is very short (about  $10^{-13}$  seconds), so the tunnel diode can be used at the highest radar frequencies.

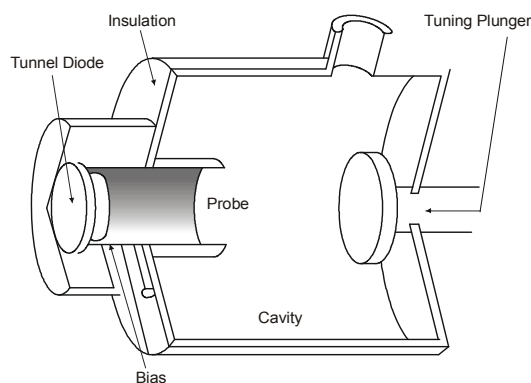
46. A typical construction of a tunnel diode is shown in cross-section in Fig 16. The p-n junction is formed by alloying a small n-type tin-arsenic bead into the surface of a p-type germanium wafer. The wafer is soldered into the casing, which forms one contact, and the n-type bead is connected to the other end of the casing via a strip of silver foil. The silver foil is soldered to the bead to act as a support as well as a contact. The bead is very small (about 0.025 mm diameter) to ensure a low junction capacitance and a high cut-off frequency.

**14-20 Fig 16 Conventional Tunnel Diode Construction**



47. **Tunnel Diode Oscillator.** Since a suitably biased tunnel diode presents a negative resistance to its terminals at all frequencies up to cut-off, it is capable of resonating at the resonant frequency of any high Q series-tuned circuit of smaller resistance connected across its terminals. It is possible to construct a resonant cavity tunnel diode oscillator capable of being tuned over a broad frequency range. The outline of a simple system is shown in Fig 17. The oscillation frequency is that of the cavity, and may be varied by altering the setting of the tuning plunger. Tunnel diode oscillators have been developed to operate satisfactorily at 100 GHz, and the cut-off frequency of available tunnel diodes is usually quoted as being in excess of 50 GHz. At this frequency, output powers of 200  $\mu$ W are possible.

**14-20 Fig 17 Resonant-cavity Tunnel Diode Tuneable Oscillator**



48. **Tunnel Diode Amplifier.** The tunnel diode is a 'majority carrier' device and is not limited in frequency by the low diffusion velocity of minority carriers. Its high frequency performance is limited only by its CR product and by the inductance of the package. By using small area junctions, and semiconductor materials with high charge-carrier mobility, the CR product may be kept low to produce cut-off frequencies in excess of 50 GHz. The useful upper frequency limit for tunnel diode amplifiers is of the order of 20 GHz. Up to this frequency, the tunnel diode amplifier can provide gains as high as 15 dB at noise figures as low as 3.5 dB. Other advantages of the tunnel diode amplifier are its reliability, and its small size and weight. By using strip-line components, very compact RF heads for radar receivers operating at frequencies around 10 GHz have been developed.

## P-I-N Diode

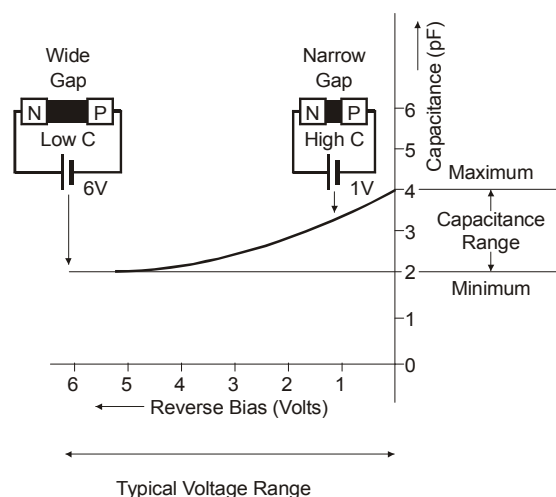
49. The p-i-n diode consists of heavily doped layers of p- and n- type silicon, separated by a thin slice of high resistance undoped (intrinsic) silicon. p-i-n diodes have been designed for use as switches at frequencies up to 40 GHz, with insertion losses as low as 0.1 dB and isolation better than 30 dB. This insertion loss means that about 98% of the incident power passes to the output when 'on', and the isolation is such that only about 0.1% of the incident power reaches the output when 'off'. For some applications, high peak power handling capability and high reverse breakdown voltage characteristics are required, and p-i-n diodes capable of handling up to 10 kW peak power with breakdown voltages of 1500 volts are now available.

50. It is possible to have an array of p-i-n diodes suitably spaced along the length of a transmission line. This gives an improvement in the attenuation and control of RF energy being passed, and also improves the reliability of the system, since the failure of any one device would result in only a small reduction in overall effectiveness.

## Varactor and Step Recovery Diodes

51. In a semiconductor diode, a capacitance is formed by the opposite charges either side of the junction. When the diode is reverse biased, the charges move further apart and the junction capacitance decreases (capacitance is inversely proportional to the distance between opposite charges). Therefore, by varying the value of reverse bias voltage applied to the diode, the capacitance can be varied, as shown in Fig 18.

14-20 Fig 18 Variation of Capacitance in Varactor Diode



52. p-n junction diodes that have been specially doped to provide a wide variation of capacitance with applied reverse bias are called 'varactor diodes'. They have many uses, eg providing a voltage controlled variable capacitance for tuning tuned circuits, in parametric amplifiers, and as electronic switches. One of their main uses is in frequency multiplier circuits, where the output frequency is a multiple of the input frequency.

53. The 'step-recovery' diode is a close relative of the varactor diode, differing in the degree of doping of the n and p regions, and being particularly suitable for frequency multiplication.



## Metal Base Transistor

54. In the metal base transistor, the usual n- or p-type base region is replaced by a thin metal film of molybdenum, which has three effects:

- a. The 'energy gap' between the emitter and the base is increased, so that the electrons from an n-type emitter are injected into the base region with high energy.
- b. The base resistance is reduced.
- c. There are no minority carriers.

55. Electrons are injected over the emitter-base barrier and collected after crossing the reverse-biased base-collector junction. Control is effected by the base voltage; the current is not diffusion limited as there are no minority carriers in the base region. The emitter current is more like the thermionic emission current in a valve; the mobility of electrons through the base is high, transit time effects are negligible, and the low base resistance of the metal base transistor further improves high frequency performance.

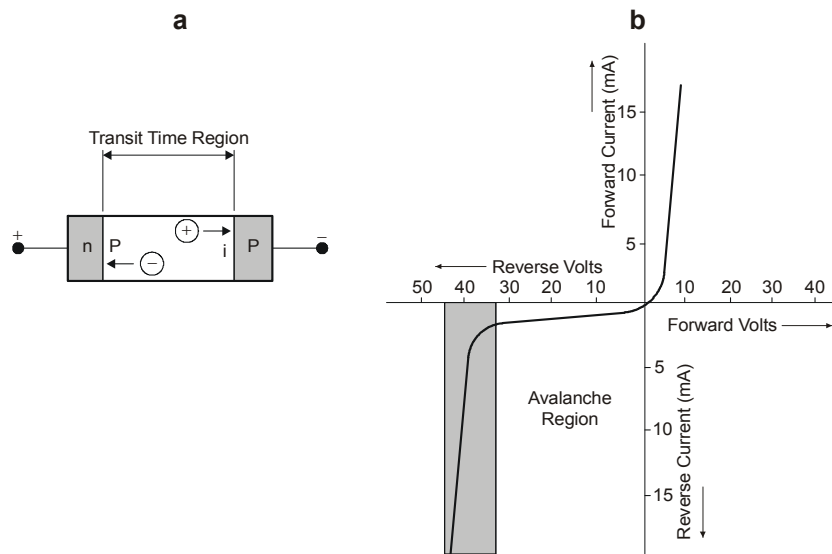
56. Metal base transistor amplifiers have been developed to give 20 dB of gain at 10 GHz, and there are indications that they will be used to provide amplifiers and oscillators in the 50 GHz to 120 GHz range.

## IMPATT Devices

57. The term IMPATT (**IMP**act **A**valanche and **T**ransit **T**ime device) is used to describe a broad class of oscillating devices which use impact ionization to create an avalanche of charge carriers which, drifting in a transit time region, produce a negative-resistance oscillation.

58. IMPATT devices include those semiconductor diodes which exhibit negative resistance characteristics, and which also depend for their operation on transit time effects. Such diodes include the silicon avalanche diode and the 'Read' diode. In general terms, an IMPATT device contains at least one semiconductor junction which is reverse biased to such a value that the resulting electric field is sufficient to spontaneously generate electron-hole pairs by a form of internal secondary emission (avalanche or multiplication process).

59. Fig 19 shows the schematic outline of a reverse biased Read diode, and illustrates a typical characteristic. The reverse bias creates a strong electric field across the transit time region. The maximum field exists at the p-n junction, and is large enough to create avalanche conditions, producing a pulse of electron-hole pairs. The generated electrons are attracted by the applied field to the nearby n-region, whilst the positive charge carriers (holes) move through the transit time region to the p-electrode. When the holes reach the p-electrode, another avalanche pulse is generated, the frequency of the pulses depending upon the transit time of the charge carriers. The phase relationship between current and voltage is such that the device exhibits negative resistance characteristics. The diode can, therefore, be used as a negative-resistance oscillator.

**14-20 Fig 19 Basic Read Diode and Characteristic**

60. The usable frequency range is partly determined by the transit time of the current carriers. Thus, for operation at 10 GHz, the transit time region would be very thin (about  $25 \times 10^{-6}$  mm). However, since the basic IMPATT device produces pulses, sinusoidal oscillations can be achieved only by mounting the diode in a microwave cavity. The same diode can be used with several different cavities to produce oscillations over a very wide frequency range. An experimental Read diode has produced oscillations in the 2 to 4 GHz band in a coaxial system, in the 7 to 12 GHz band in an I-band waveguide, and at 50 GHz in a millimetric waveguide.

61. A silicon avalanche diode has been developed to provide outputs of 1W CW at 12 GHz with 8% efficiency, and a few milliwatts CW at 50 GHz with 2% efficiency. It seems likely that these figures can be improved upon to provide perhaps 20W in the 5 to 20 GHz range, with efficiencies of some 30%. As an amplifier, IMPATT devices have produced gains of 20 dB at 10 GHz with 30 MHz bandwidth, although the high noise of 50 dB prohibits its use as an RF amplifier.

### TRAPATT (Trapped Plasma Avalanche Transit Time) Diode

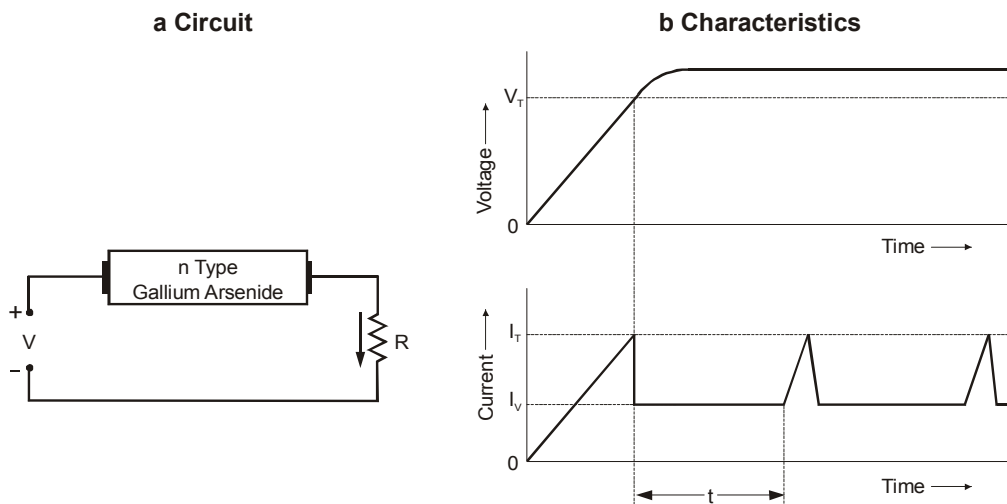
62. Whilst trying to improve the efficiency of the IMPATT by experimenting with the diode/cavity geometry, it was accidentally discovered that an anomalous mode of operation is possible. If the diode is installed in a cavity whose resonant frequency is approximately half of the IMPATT operating frequency, and a voltage approximately twice the breakdown voltage is applied, unusually high power and high efficiency can be achieved. The overvoltage produces an avalanche shock front, which propagates through the diode at a velocity that is greater than the velocity of the charge carriers. Behind this front, a high-density plasma is trapped in the low field region that has been created. These trapped carriers then move relatively slowly across the diode and maintain a high current in the external circuit, whilst the applied voltage is low. Diodes operating in the TRAPATT mode can achieve up to 20 W pulsed at an efficiency of up to 30%. Although they are currently limited in operating frequency to about 16 GHz, this is quite high enough for airborne applications, such as radar transponders and TACAN equipment.

### Gunn Effect Oscillator

63. When the voltage applied to a thin slice of gallium arsenide semiconductor exceeds a critical threshold value, periodic fluctuations of current result and, if the slice is thin enough ( $4$  to  $25 \times 10^{-6}$  mm), the frequency of oscillation is in the microwave band. The effect is known as the 'Gunn' effect, after its discoverer. A Gunn effect semiconductor is called a 'bulk' device since there are no junctions.

64. Fig 20a shows a basic circuit of a gallium arsenide slice connected, in series with a small resistor, to a source of dc voltage. If the current through the resistor is measured, it is found that it is proportional to voltage up to a certain critical threshold value of voltage ( $V_T$ ). However, when the voltage is increased above  $V_T$ , the current drops rapidly from its value ( $I_T$ ) to a valley current value ( $I_V$ ). The current then remains at the value  $I_V$  for a time ( $t$ ) before rising to its original value ( $I_T$ ) (Fig 20b). The period of oscillation is related to the time ( $t$ ), which depends upon the slice thickness, but is independent of the applied voltage.

**14-20 Fig 20 Gunn Effect in Gallium Arsenide Semiconductor**



65. Gunn oscillators can be tuned to operate over a band of frequencies by mounting the diode in a microwave resonant cavity of variable length. Operated in a pulse mode, the output power ranges from 200 W at 1 GHz to 1 W at 10 GHz. CW outputs between tens and hundreds of mW are available. These devices are simple and reliable, and find applications in receivers, as local oscillators, and in low-power transmitters.

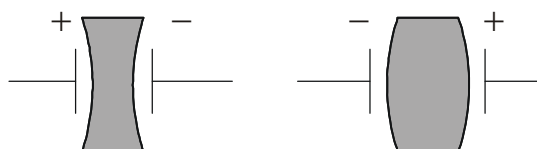
## **SURFACE ACOUSTIC WAVE (SAW) DEVICES**

### **Introduction**

66. Surface acoustic wave devices are becoming increasingly important as components in radar systems where, typically, they may be used as delay lines, pulse generators, and to produce frequency modulated pulse compression pulses. The underlying phenomenon is that of the piezo-electric effect.

67. **Piezo-Electric Effect.** The piezo-electric effect is exhibited by a number of crystals, such as quartz. It manifests itself as a change in physical shape in response to the application of a voltage across a slab of material made from such crystals. In effect, the material expands or contracts depending on the polarity of the voltage, as shown in Fig 21. Therefore, if an oscillating voltage is applied, the material will vibrate at the frequency of the applied voltage. The effect is reversible; thus, if a piezo-electric crystal is alternately compressed and expanded, an oscillating voltage will appear across its face.

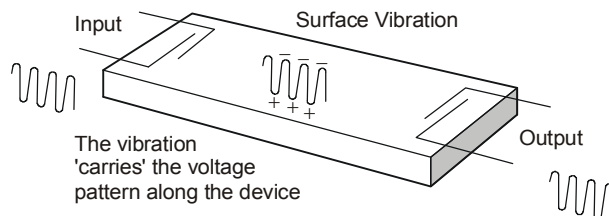
**14-20 Fig 21 Piezo-Electric Effect**



### Basic SAW Device Construction

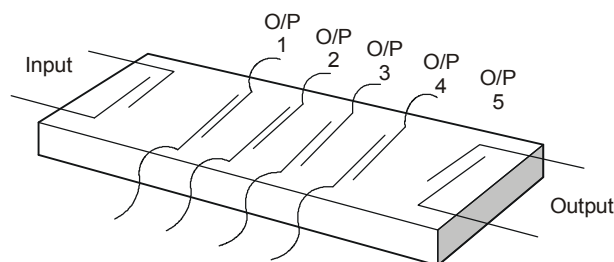
68. A basic SAW device consists of a slab of piezo-electric crystal material with metal inter-digital fingers deposited on one surface, as shown in Fig 22. When an input oscillating signal is applied, a voltage is created across the input finger. This produces a mechanical vibration in the surface material. This physical vibration travels along the crystal slab, mimicking the electrical input signal and, as the vibrations pass the output fingers, an oscillating voltage is generated.

14-20 Fig 22 Basic SAW Device



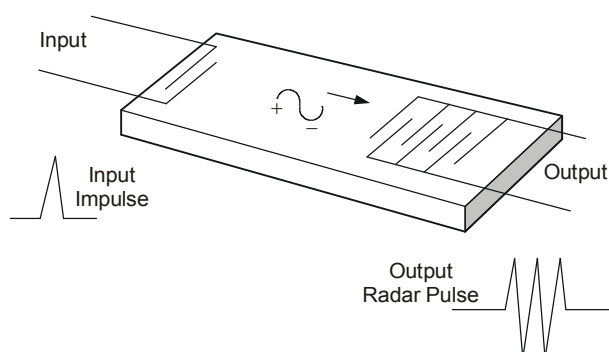
69. **Delay Line.** The speed at which the mechanical vibration travels in the slab is some  $10^5$  times slower than the speed of electromagnetic waves in space; so, the simple device of Fig 22 could act as a delay line. For a  $1\mu\text{s}$  delay, the device need only be 3 mm long, compared to 100 ft of cable to achieve the same effect. The delay can be varied by placing a series of inter-digital fingers at different positions along the length of the device, so that the output can be tapped after any desired delay (Fig 23).

14-20 Fig 23 Tapped Delay Line



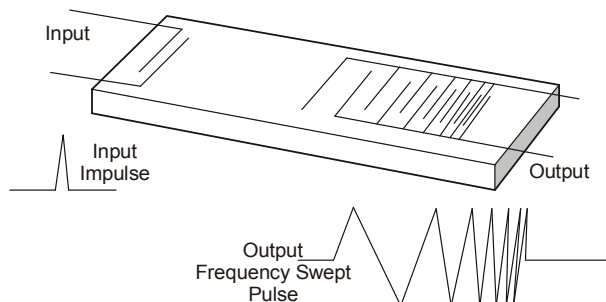
70. **Pulse Generation.** If a single voltage spike is applied at the input of the simple device of Fig 22, the output will also be a single spike. However, if the output contacts are modified to be a series of inter-digital fingers (Fig 24), an oscillating pulse is produced. As the mechanical vibration pattern passes the first two fingers, a positive spike is generated and, as it passes the second pair, a negative spike is produced, and so on. The number of cycles in the pulse depends on the number of inter-digital fingers at the output, and the wavelength on the spacing of the fingers.

14-20 Fig 24 SAW Pulse Coding

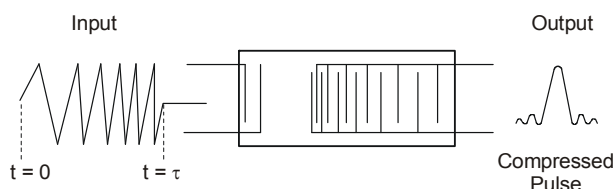


71. **Production of an FM Pulse Compression Pulse.** The device illustrated in Fig 24 can be modified to produce an output pulse with a varying frequency by varying the spacing of the output inter-digital fingers, as shown in Fig 25. By reversing the device, it can act as a detector for such a frequency modulated pulse. In Fig 26, the FM pulse is applied to the input. The resulting surface wave will have an identical pattern, and will only produce a large output when the pattern exactly overlaps the output contacts, ie the device compresses the pulse into a narrow width.

**14-20 Fig 25 FM SAW Device**



**14-20 Fig 26 A SAW Device for Pulse Compression**



## LOW NOISE AMPLIFIERS

### Introduction

72. If there was no noise present in a receiver, it would be possible to detect any signal, given sufficient amplification. The noise generated within a practical receiver, compared with a 'noiseless' receiver, is expressed by the noise figure (F) where:

$$F = \frac{S_{in}/N_{in}}{S_{out}/N_{out}}$$

where  $S_{in}$  = input signal power

$S_{out}$  = output signal power

$N_{in}$  = input noise power

$N_{out}$  = output noise power

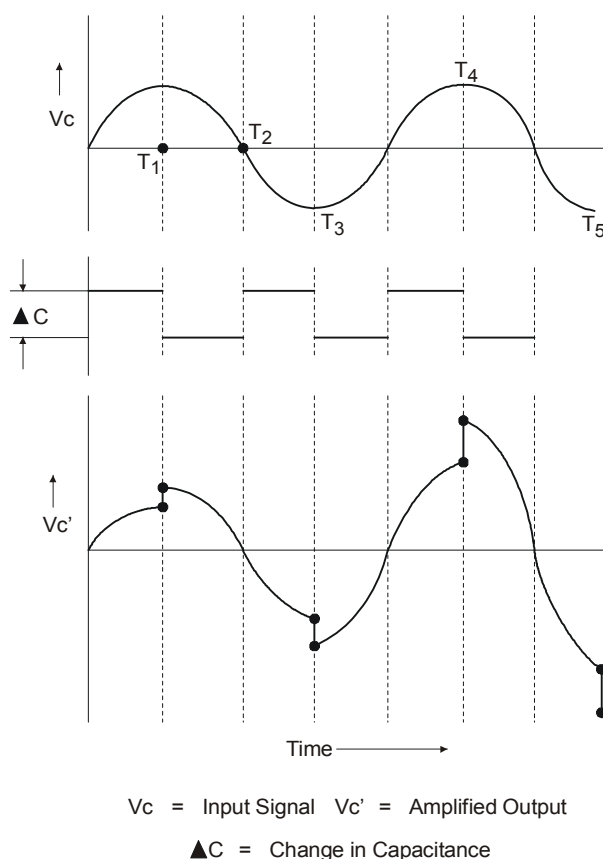
73. Amplifiers in a superhet receiver generate sufficient noise to negate the gain they provide. However, low noise amplifiers have been developed, such as:

- The travelling-wave tube (paras 18 - 24).
- The parametric amplifier.
- The maser.

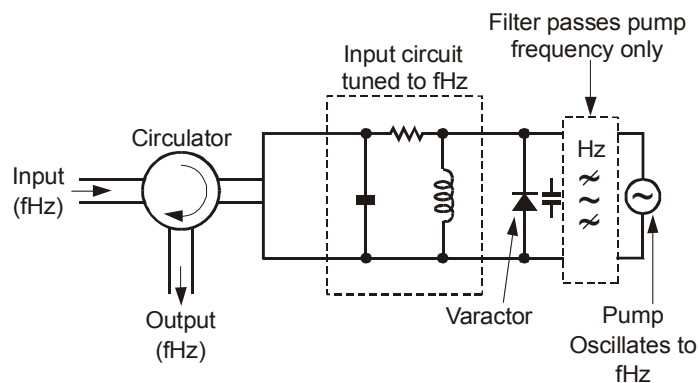
### The Parametric Amplifier

74. The parametric (or reactance) amplifier derives its name from the fact that the equation governing its operation has one or more parameters that vary with time. The principle of operation may be explained by considering a simple resonant circuit having an inductance and a capacitance oscillating at the resonant frequency. If the capacitor plates were to be pulled apart at the instant when the oscillating voltage was at a positive maximum, then, since for a fixed charge the voltage is inversely proportional to the capacitance, the voltage would increase. If the plates were returned to their original position as the voltage passed through zero, then, since there would be no charge on the plates, the voltage would remain unchanged. If this process was repeated, separating the plates every time the voltage passed through a positive or negative maximum, and returning them to the original position as the voltage passed through zero, then a signal at the resonant frequency would be amplified. The principle is illustrated in Fig 27 where, at time  $T_1$ , corresponding to a voltage maximum, the capacitance decreases, leading to a step increase in voltage. There is no change at  $T_2$ , where the voltage is zero, but there is a further step at  $T_3$ , where the voltage is at a negative maximum. The process is repeated at  $T_4$ ,  $T_5$  and so on. The variable capacitor element of a parametric amplifier is normally provided by a varactor diode (see para 51), in which the depletion layer at the junction may be considered as the dielectric between the plates of a capacitor. Under reverse bias, the width of the depletion layer varies with the applied electric field and hence, if an oscillating voltage is applied across a varactor diode, the capacitance will vary at the oscillating frequency.

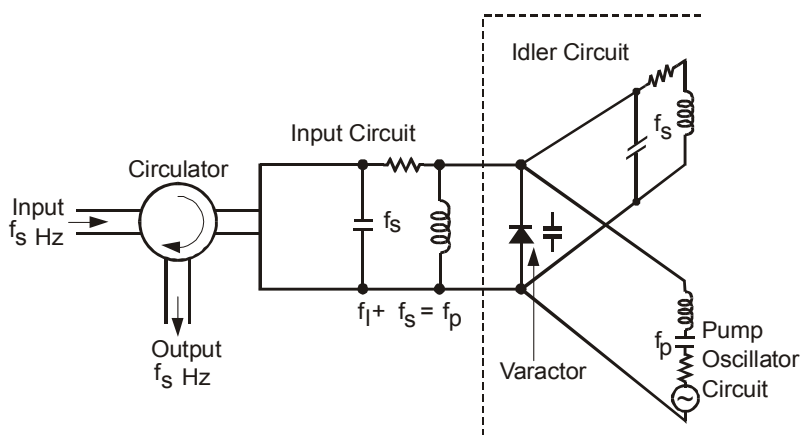
14-20 Fig 27 Principle of Parametric Amplifier



75. **Degenerate Parametric Amplifier.** The simple parametric amplifier described above requires that the variation in the capacitance occurs at a frequency, called the 'pump' frequency, which is twice that of the resonant frequency of the resonant circuit. This mode of operation is known as the 'degenerate mode', and amplification takes place only if the phase relationship between the pump frequency and the signal frequency is correct. The essentials of a basic degenerate parametric amplifier are shown in Fig 28.

**14-20 Fig 28 Degenerate Parametric Amplifier**

**76. Non-degenerate Parametric Amplifier.** The phase sensitivity of the degenerate parametric amplifier can be overcome by operating the pump at some frequency other than twice the resonant frequency. The nature of the parametric amplifier is such that a third frequency, known as the 'idling frequency', is produced. The idling frequency is equal to the difference between the pump frequency and the signal frequency. The phase and frequency limitations can be relaxed if an idler circuit, resonant at the idling frequency, is added, as shown in Fig 29. In effect, this idler circuit acts as an energy reservoir, accepting energy from the pump or signal circuits, and storing it until needed, then releasing it at the proper time and phase, to provide power gain in the signal circuit. The non-degenerate parametric amplifier, in which  $f_p = f_s + f_i$  (frequencies of pump, signal, and idler circuits respectively), is a negative resistance device; it has limited gain and, like any other negative resistance amplifier, has a tendency towards instability.

**14-20 Fig 29 Non-degenerate Parametric Amplifier**

**77. Parametric Up-converter.** The problem of instability can be solved by a parametric amplifier which presents a positive resistance to the signal circuit, by making  $f_p = f_i - f_s$ . This is called an 'up-converter', and is completely stable. The output is at the idling frequency, which is higher than the signal frequency, and this amplifier can be followed by a conventional crystal mixer receiver. The up-converter has a maximum gain which is proportional to  $f_i/f_s$ , so is primarily useful at UHF frequencies or lower.

**78. Other Parametric Devices.** If  $f_p = f_s - f_i$ , the device becomes a 'positive resistance down-converter'. It has a gain of less than unity, and is used as a mixer in preference to the crystal diode. Negative resistance up- and down-converters are also available; these have high gains but have tendencies towards instability.

**79. Properties of the Parametric Amplifier.** The parametric amplifier has a very low noise figure, typically about 2 to 3 dB at room temperature; this can be improved further by refrigeration using, for example, liquid nitrogen at 77 K. This low noise is mainly the result of not employing a noisy electron beam generated by a hot cathode. A parametric amplifier can be used as a signal frequency amplifier in radar receivers to improve the receiver sensitivity, and hence increase the range of the radar. It is mounted between the TR cell and the crystal mixer and is a 'fail safe' device, in that failure of the parametric amplifier does not result in failure of the radar, although there will be a decrease in signal-to-noise ratio. The noise present in a parametric amplifier is due to the noise introduced by the varactor and the thermal noise generated in leads, circulators, waveguide walls etc; the noise figure increases with increasing operating frequency. The gain of a parametric amplifier depends on the pump power supplied, and the pump frequency, and is limited by possible instability. The output from a klystron pump oscillator is often fed to the varactor through an attenuator in order to give the required gain consistent with stability. Pump power is normally a few mW. For a given input frequency, a high gain can be obtained by using a high pump frequency; however, it is difficult to generate useful power at very short wavelengths. A typical gain is 20 dB, and, to obtain this with an input frequency of 600 MHz, 20 mW of pump power is used at a frequency of 8.9 GHz.

## Masers

**80.** The maser is an extremely low noise amplifier, based on quantum mechanical principles, and capable of noise temperatures in the order of a few degrees Kelvin. The name is an acronym formed from **M**icrowave **A**mplification by **S**timulated **E**mission of **R**adiation. A maser requires complicated external circuitry and a magnetic field and, furthermore, it has a narrow bandwidth and must be operated at liquid helium temperatures (about 1 to 4 K). The principle of operation of the maser is identical to that of the laser, in that the electrons in the atoms of a material are pumped to a higher than normal energy level from which they are stimulated to jump back to a lower level with the emission of energy. The frequency of the emission is dependent on the difference in energy levels and is given by:

$$f = \frac{\text{Difference in energy levels}}{h \text{ (Planck's constant)}}$$

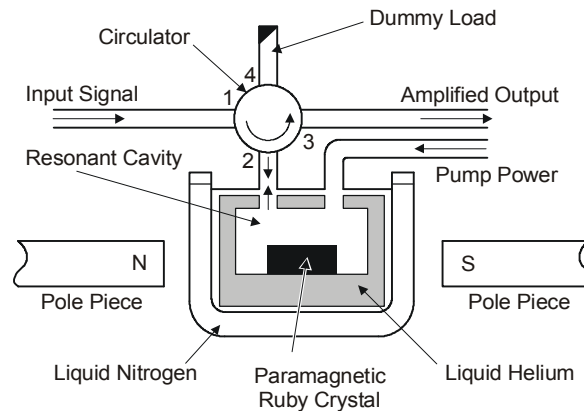
There are several forms of maser and those of chief interest in radar applications are solid-state devices using paramagnetic crystals placed in magnetic fields.

**81. Three-level Cavity Maser.** The three-level cavity maser, in which a ruby maser material has three closely spaced energy states ( $E_1$ ,  $E_2$ , and  $E_3$ ), is illustrated in Fig 30.  $E_2$  and  $E_3$  are separated by an energy difference which would lead to radiation at the frequency of the signal requiring amplification. A strong RF pump signal is applied to the system at a frequency of  $f_p = (E_3 - E_2)/h$ . By exciting electrons into higher energy levels, this forces atoms from the lower energy state  $E_1$  to the higher state  $E_3$  until the populations in the two levels are equal. There are now more atoms in  $E_3$  than in  $E_2$ , and the application of the signal which requires amplification ( $f_s = (E_3 - E_2)/h$ ) acts as a stimulus to cause atoms to drop to the  $E_2$  level, with an accompanying emission of radiation at frequency  $f_s$ , thereby amplifying the signal. The separation of the energy states depends upon the strong magnetic field in which the paramagnetic material is placed. Thermal processes can, however, mask this maser effect, hence the need to cool the material to liquid helium temperature. Low noise is an important outcome of this low temperature operation. Since the cavity maser is a one-port device, the amplified signal must be taken from the cavity via the same line as it entered, and a circulator is used to separate the input signal from the amplified output signal. A typical three-level cavity maser, operating at 1.5 K, amplifies a signal of 2.8 GHz with a gain of 20 dB. The pump power is about 10 mW at 9.4 GHz, the



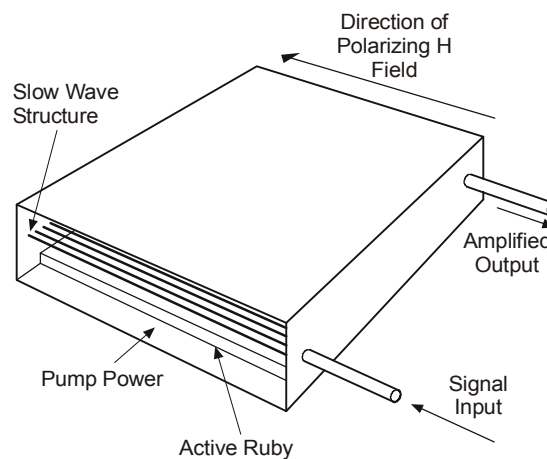
bandwidth 20 MHz, and noise figure about 0.3 dB. An I-band maser amplifying a signal of 9.4 GHz, requires a pump frequency of 24 GHz. The disadvantages of this type of maser are its tendency to oscillate and its narrow bandwidth, both of which are overcome in the travelling wave maser.

**14-20 Fig 30 Three Level Cavity Maser**



**82. Travelling Wave Maser.** The use of a travelling wave structure, rather than a cavity, allows bandwidths to be improved to a reasonable figure. The basic arrangement, shown in Fig 31, consists of a ruby loaded comb structure mounted in a waveguide. Pump power is fed down the waveguide and this causes a population inversion in the ruby crystal. The input signal to the slow-wave structure causes stimulated emission, and is thus amplified. A magnetic field, in the direction shown, is provided and the whole structure is immersed in a liquid helium bath. A typical travelling wave maser, operating at 19 GHz, has a gain of 20 dB, a bandwidth of 55 MHz, and a noise figure of 0.16 dB.

**14-20 Fig 31 Travelling-wave Maser**



## WAVEGUIDES

### Introduction

**83.** Above about 1 GHz, dielectric and skin losses in coaxial cables become so great that only very short lengths may be used. The greatest contribution to the loss comes from the inner conductor, and the solution to the problem is to remove it. Thus, a waveguide is essentially a hollow air-filled pipe made of a conducting material; it may be considered as a medium of one dielectric constant (air) enclosed in another medium of different dielectric constant (metal tube). In theory, the cross-section of the waveguide may take any form but, in practice, it is usually rectangular, although circular

waveguides are used for some specialized applications. The dimensions of the cross-section are dependent on the frequency of the wave which is to be transmitted along the waveguide and, as a general rule, the dimensions decrease as the operating frequency increases.

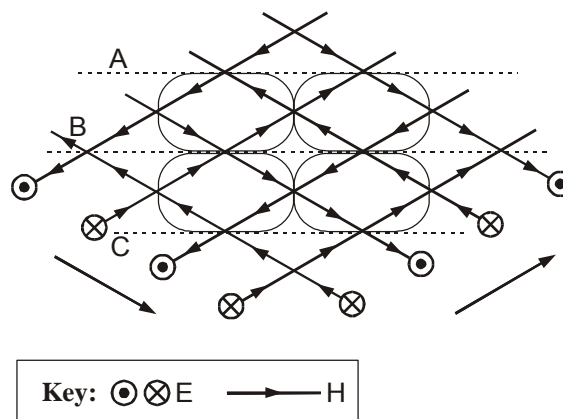
### Propagation in Waveguides

84. Where an electromagnetic wave passes close to a conducting surface, certain boundary conditions must be fulfilled. These are that the electric field (E) must be normal to the conducting surface, or zero, and the magnetic field (H) must be parallel to the conducting surface, or zero. If an attempt is made to pass a transverse electromagnetic (TEM) wave straight down the axis of a rectangular waveguide, the boundary conditions will not be satisfied at one pair of the conducting surfaces. Thus, a TEM wave is not capable of being propagated straight down the waveguide.

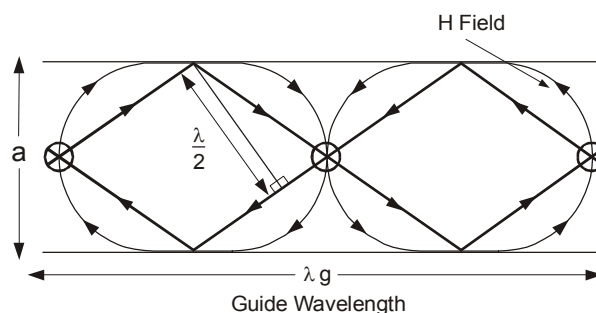
85. If a TEM is passed into a waveguide, such that it satisfies the boundary conditions at the broad faces, but is travelling at an angle to the axis of the waveguide, it will be reflected by the narrow faces (at which the boundary conditions will also be met).

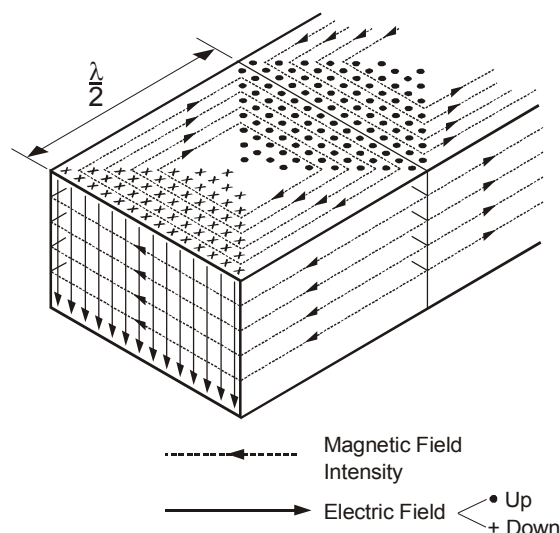
86. Fig 32 shows how two such waves, at an angle to each other, would interact. Along the planes A, B, and C, perpendicular to the page, the resultant E field is zero, and the H field between these planes forms loops as shown. Thus, conducting plates could be placed along these planes and the boundary conditions would be satisfied. A further pair of plates can be added to complete a rectangular waveguide, and these too will satisfy the boundary conditions. Fig 33 shows the resulting pattern which, since the fields change polarity at the signal frequency, propagates along the waveguide. The wavelength of the pattern, known as the 'guide wavelength' ( $\lambda_g$ ), is longer than the free space wavelength. Fig 34 shows the arrangement in three dimensions.

14-20 Fig 32 Interaction of Two TEM Waves



14-20 Fig 33 Field Pattern Resulting from Two TEM Waves



14-20 Fig 34 Three-dimensional Illustration of H<sub>10</sub> Mode

87. The pattern described is known as the H<sub>10</sub> mode; other modes are possible but the H<sub>10</sub> is the most efficient. The particular mode which is established depends upon the dimensions of the waveguide but, if the broad dimension of the waveguide ( $a$ ) lies between  $\lambda$  (the free space wavelength) and  $\lambda/2$ , only the H<sub>10</sub> mode will be propagated. If  $a = \lambda/2$ , a wave will not propagate, and there is, therefore, a critical wavelength ( $\lambda_c = 2a$ ) corresponding to a cut-off frequency,  $f_c = \frac{3 \times 10^8}{\lambda_c}$ , below which the waveguide will not work.

88. The next highest mode of propagation is established if  $\lambda = a$  or  $2b$  (where  $b$  is the narrow waveguide dimension), whichever is the greater. In order to prevent arcing, and to allow maximum power handling,  $b$  should be as large as possible and is normally made equal to  $a/2$ . In summary, for only H<sub>10</sub> propagation:

$$a \text{ (or } 2b) < \lambda < 2a, \text{ or } \frac{3 \times 10^8}{a} > f > \frac{3 \times 10^8}{2a}$$

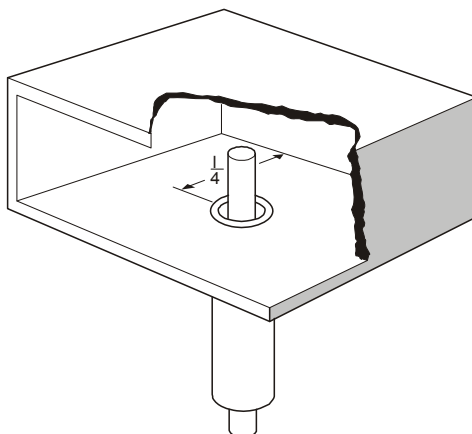
89. **Power Carrying Capacity.** There is a limit to the power which can be carried by any transmission line. In an air-filled waveguide, there is no loss in the air and the power dissipated in the walls is seldom a limiting factor. The limit is set by voltage breakdown in the air. Dry air, at atmospheric pressure, breaks down when the electric field strength reaches about 30 KV/cm. For the practical power rating of waveguides, it is usual to take the safe field strength limit as 15 KV/cm. Since the dimensions of a waveguide are a function of wavelength, their power-carrying capacity decreases as frequency increases. Changes in pressure also affect breakdown, and it is found that the breakdown voltage is approximately proportional to pressure over a wide range. This presents an obvious limitation for airborne equipment. For example, at 35,000 ft, where the air pressure is approximately a quarter of that at ground level, the power carrying capacity of the waveguide is reduced by a factor of 16 (since power is proportional to the square of voltage). This effect can be obviated by sealing off the waveguide system so that the interior is maintained at, or near, standard atmospheric pressure. Conversely, the power-carrying capacity can be increased above the normal by increasing the pressure inside the guide above standard pressure.

### Waveguide Techniques and Components

90. **Launching the Wave.** Two methods are commonly employed to launch the waves into the guide. The first method employs a suitable probe, connected to a short run of coaxial line, which is fed by the source of the waves. In its simplest form, the probe is merely the inner conductor of the

coaxial line, extended across the narrow dimension of the guide, the outer conductor of the line being connected to the waveguide itself (Fig 35). The inner conductor acts as a small aerial and radiates the required component TEM waves of the waveguide mode. The waveguide short circuit is used to reflect the waves in the required direction, and is placed  $\lambda_g/4$  behind the probe, so that the correct phase relationship is obtained. The other method of launching the wave is by connection to a resonant cavity by means of a slot.

**14-20 Fig 35 Launching from a Coaxial Line by a Simple Probe**



91. **Extracting the Wave.** When a waveguide is used to connect an aerial to a receiver, the wave is often extracted by means of a probe, as used for launching the wave. When the waveguide is connecting a transmitter to an aerial, the method of extraction is usually one of the following:

a. **A Matching Horn.** If the waveguide terminated in a simple open end, some energy would leak out, but some would also be reflected. To prevent the reflection, the open end of the waveguide is flared out to form a horn, effectively matching the impedance of the guide to that of free space; the impedance of the guide gradually changing over the flared section.

b. **Coupling Slots.** The H field in a waveguide causes currents to flow in the inside walls, these currents being at right angles to the H field. If a slot is cut in the wall of the waveguide, such that it interrupts the flow of wall current, it will radiate energy. Fig 36a shows slots cut in the broad and narrow sides of a waveguide. Slots 1 and 2 will interrupt the wall current at right angles, so maximum coupling occurs; slots 3 and 4 are parallel to the wall current and no energy is coupled; slot 5 is inclined to the wall current and slot 6 is displaced from the centre of the broad side. As both slots 5 and 6 interrupt the current, they will couple energy from the guide, but with reduced effectiveness compared to slots 1 and 2. For least disruption of the waveguide field pattern, the slots should be half a wavelength long. In practice, a series of half-wave slots is cut in the wall of the waveguide, arranged so that they each radiate an equal amount of in-phase energy. Two systems used are:

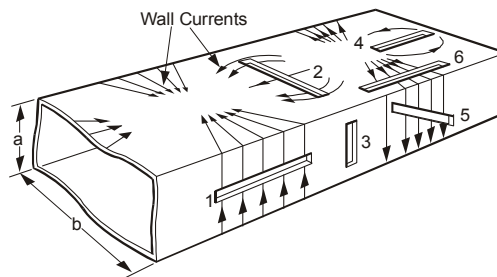
(1) **Displaced Slots.** Displaced slots are slots of length  $\lambda_g/2$  cut parallel to, but not along, the centre line of the broad side (e.g. slot 6 in Fig 36a). As the wall current reverses phase every half a guide wavelength, the slots are spaced with their mid-points  $\lambda_g/2$  apart on alternate sides of the centre line, successive slots being further removed from the centre line to give equal radiation from each slot (Fig 36b).

(2) **Inclined Slots.** Inclined slots are slots of length  $\lambda_g/2$  cut at an angle in the narrow side of a waveguide, e.g. slot 5 in Fig 36a. As the angle of inclination ( $\theta$  in Fig 36c) is increased, the

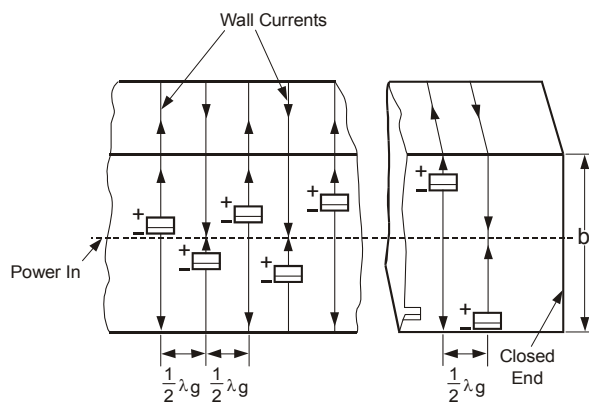
coupling increases. With an array of slots, the angle of inclination is progressively increased to compensate for the reduction of energy in the guide, so that each slot radiates equal energy. If the slots were spaced  $\lambda_g/2$  apart, and inclined in the same direction, adjacent slots would radiate in anti-phase. By alternating the slopes of adjacent slots, as in Fig 36c, the slots radiate in phase, and a beam at  $90^\circ$  to the line of the array is formed. As the slots are half a wavelength long, they continue into the broad side of the waveguide.

**14-20 Fig 36 Coupling Slots**

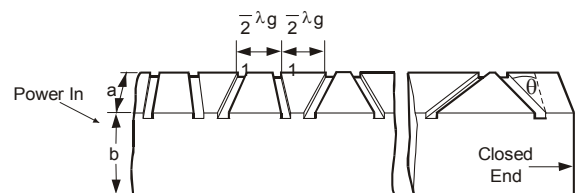
**a Slot Positions in Relation to Wall Current**



**b Displaced Slots**

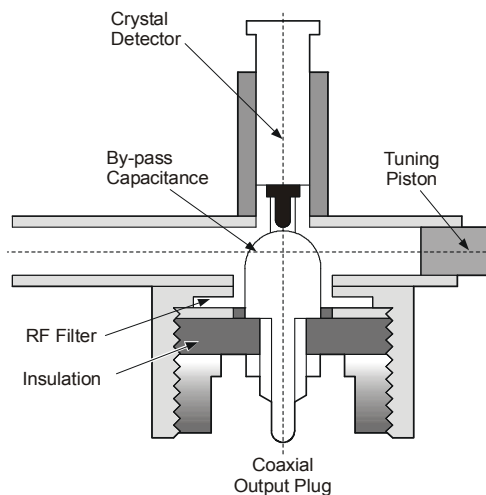


**c Inclined Slots**



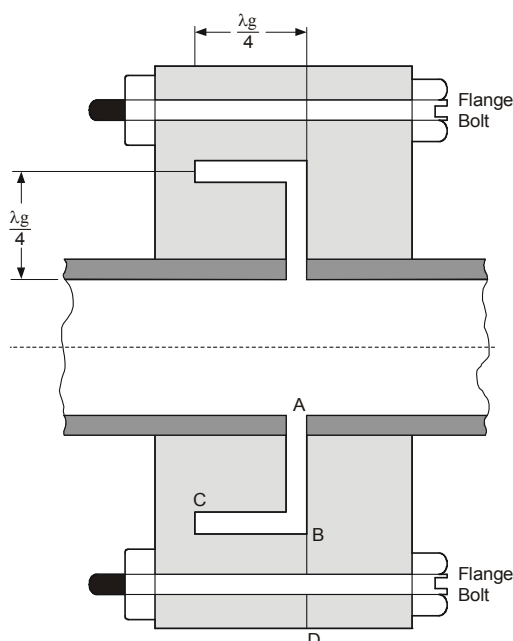
**92. Detecting the Wave.** To detect signals travelling down a waveguide (e.g. for test purposes), it is usual to mount the detector directly on, or in, the waveguide, the physical size of the normal crystal detectors being such that this is readily possible. Fig 37 shows a coaxial-type crystal, mounted on a 'doorknob' transformer, acting as a detector in a rectangular waveguide.

**14-20 Fig 37 Crystal Detector Mounted on a Waveguide**

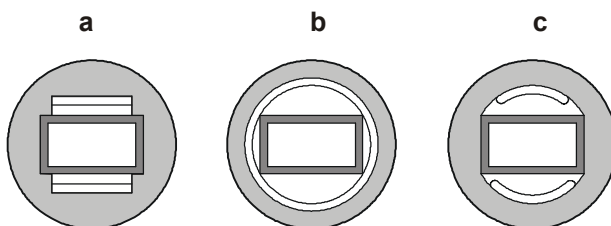


93. **Choke Joints.** Problems with installation and maintenance demand that waveguide circuits be made in a number of rigid sections which must then be joined together. Early systems used two plain flanges on the ends of the pieces to be joined, which were then clamped mechanically in order to obtain good electrical continuity. A better system is that used in the resonant, short-circuit, choke joint, shown in section in Fig 38 and end-on in Fig 39a. One flange is still plain, but the other is machined to form an effective half wavelength of transmission line when bolted to the plain flange. Due to standing waves, the short circuit at C sets up an effective short circuit at A, on the waveguide wall, thus giving good continuity. Since the point B is at a region of high impedance (and hence low current), poor contact between the coupled sections will not have any appreciable effect. It should be noted that good continuity is not required on the narrow faces of the guide since the electric field there is zero. For this reason, the choke BC is often machined in circular fashion, as shown in Figs 39b and c.

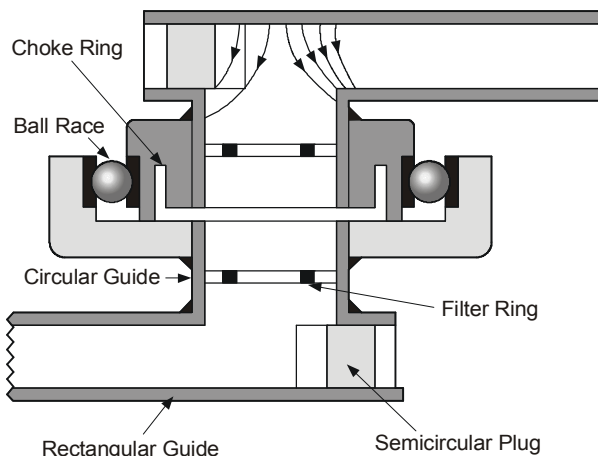
14-20 Fig 38 Section of Choke Coupling



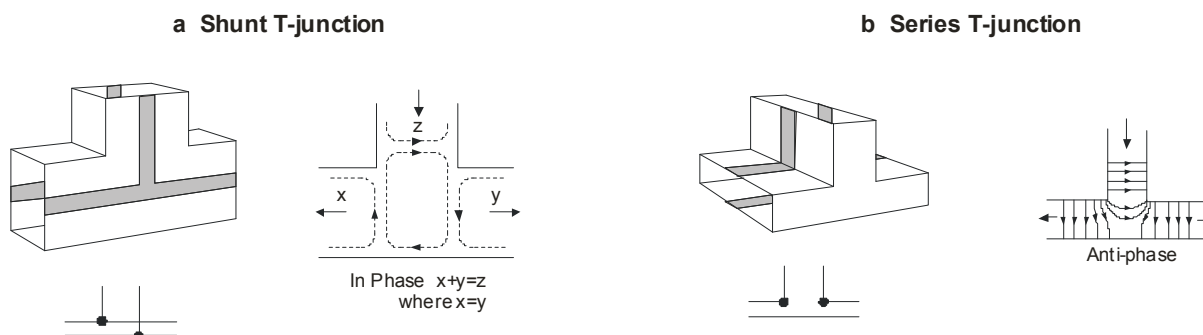
14-20 Fig 39 Various Types of Choke Flanges



94. **Rotating Joints.** In many radar and communications equipments, it is necessary to rotate the aerial with respect to the rest of the equipment and, for this purpose, a rotating joint must be provided in the waveguide feed. A section through a typical rotating joint is shown in Fig 40. A circular waveguide is used for the rotating section. The wave travels along the circular guide in circular mode, which has a radially symmetrical field pattern. The input to, and output from, the joint is by rectangular guide. Transformation from the rectangular mode to the dominant circular mode is assisted by the semicircular plugs in the stub ends of the rectangular guide. Unfortunately, this dominant mode does not possess circular symmetry. To suppress it, ring filters are fitted above and below the choke joint, producing a field which possesses circular symmetry. This allows rotation of one part of the joint relative to the other, without disruption of field patterns.

**14-20 Fig 40 Waveguide Rotating Joint**

95. **T-junction.** There is frequently a necessity to split the energy carried in a waveguide into two portions, or combine the energy carried in two waveguides into a single guide. This is usually carried out by T-junctions, the two basic forms of which, together with their transmission line equivalents and field patterns, are shown in Fig 41. If the junction is set into the narrow face of the waveguide, it is an H-plane or shunt junction, whereas, if it is made in the broad face of the guide, it is an E-plane or series junction. The two types of junction are similar in that, if power is fed into the vertical arm of the junction, it is split evenly between the other two arms of the junction. However, they differ in that the resultant waves are in phase for the shunt junction, but in anti-phase for the series junction.

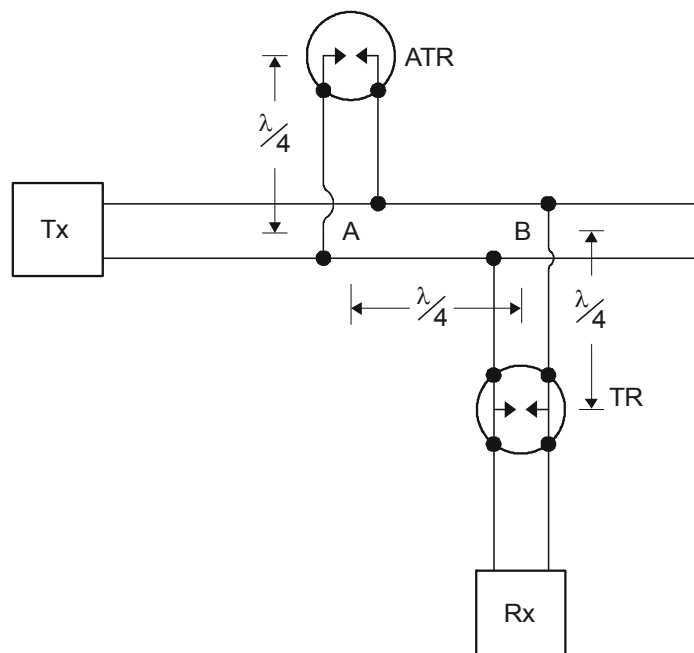
**14-20 Fig 41 T-junctions**

96. **TR Switches.** In a pulse radar system which uses the same aerial for transmission and reception, it is necessary to include in the waveguide system a transmit-receive (TR) switch (also known as a 'Duplexer'). This automatically couples the transmitter to the aerial for the duration of each pulse, while at the same time protecting the receiver from burn-out or damage. It also couples the aerial to the receiver during the receiving period. The most common types are:

- a. **Branched Duplexers.** The principle of the branched duplexer is illustrated in Fig 42. It consists of a TR cell and an anti-transmit-receive (ATR) cell placed on the waveguide (or transmission line), as shown. The ATR cell is properly termed a 'transmitter blocker' (TB) cell. Both of the cells are special gas-discharge valves, such as the soft rhumbatron, which ionize and provide a near short-circuit when the transmitter is turned on. Since both cells are situated  $\lambda_g/4$  from the main feeder in shunt arms, the short-circuits at the cells are transformed to very high impedances across the main feeder at points A and B. Thus, none of the transmitter power flows to the receiver. When the transmitter pulse is ended, the switches rapidly de-ionize and offer a high impedance. This is transformed by the

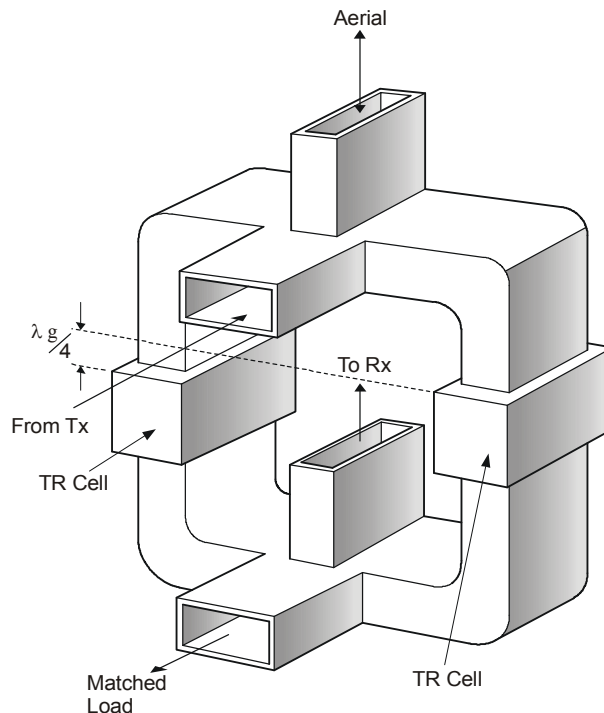
$\lambda_g/4$  sections to represent low impedances at points A and B. The low impedance at A is further transformed by the  $\lambda_g/4$  section A-B to a high impedance at B. Received signals therefore pass down the receiver branch and do not pass B towards the transmitter.

**14-20 Fig 42 Branched Duplexer Principles**



b. **Balanced Duplexers.** There are many circuit arrangements of TR and ATR cells for switching. The branched duplexer is probably the simplest configuration, but it is not inherently broadband. A more broadband arrangement is the balanced duplexer whose bandwidth is limited only by that of the waveguide. A combination of series and shunt T-junctions (see Fig 41) is known as a 'hybrid T-junction' (sometimes called a magic T-junction). Hybrid T-junctions are used in conjunction with the TR cells, and a common arrangement is shown in Fig 43. The transmitter power enters the magic-T by the arm shown, and splits equally left and right in phase. The two signals cause the TR cells, which should be identical and acquired in matched pairs, to ionize, thus acting as short circuits. The two signals are totally reflected, but since their path lengths differ by  $\lambda_g/2$  by the time they arrive back at the top magic-T, they leave by the arm which is connected to the aerial. If any small fraction of the transmitter power manages to penetrate the cells, two signals arrive at the bottom magic-T in phase and leave by the shunt arm, which is connected to a matched load where they are dissipated. The receiver is thus completely isolated from the transmitter power. On receive, the signal arriving from the aerial splits equally left and right in anti-phase, and none of it enters the transmitter. The two signals are not sufficiently strong to ionize the TR cells and so pass straight through. They arrive at the bottom magic-T, still in anti-phase, and so combine to leave by the arm connected to the receiver. In addition to this arrangement, a low power TR cell is often placed between the balanced duplexer and the receiver to safeguard the receiver against random pulses from nearby radar equipments. These pulses may be too weak to ionize the cells, but could be strong enough to damage the receiver.



**14-20 Fig 43 Balanced Duplexer**

97. **Miscellaneous Waveguide Components.** The foregoing list of waveguide components is by no means exhaustive and the scope of this chapter will permit only a brief mention of the following components:

- a. **Bends and Twists.** Often, a change of direction is necessary in a waveguide run. This can be done simply by the use of a length of guide, bent as required. The overriding principle in such a device is that any change must be made gradually to avoid reflections.
- b. **Attenuators.** In making measurements on waveguide components, and for other purposes, it is often desirable to absorb some or all of the incident energy. This can be achieved by the use of attenuators, of which there are several types.
- c. **Matching Devices.** Most waveguide components introduce a certain amount of mismatch, which has to be removed by some matching device. Matching can be carried out by means of waveguide stubs of adjustable length, in the same way as matching stubs are used with two-conductor transmission lines, or by means of lumped reactances, known as 'irises', placed in the waveguide.
- d. **Components for Circular Polarization.** Circularly polarized waves have many radar applications (eg in discriminating against rain). One of the most popular devices for producing such waves is the dielectric quarter-wave plate, which consists of a longitudinal dielectric slab in a section of circular waveguide, placed at  $45^\circ$  to the plane wave. This introduces the correct amount of phase delay in one component of the wave to cause the emerging wave to be circularly polarized.

## MICROWAVE AERIALS

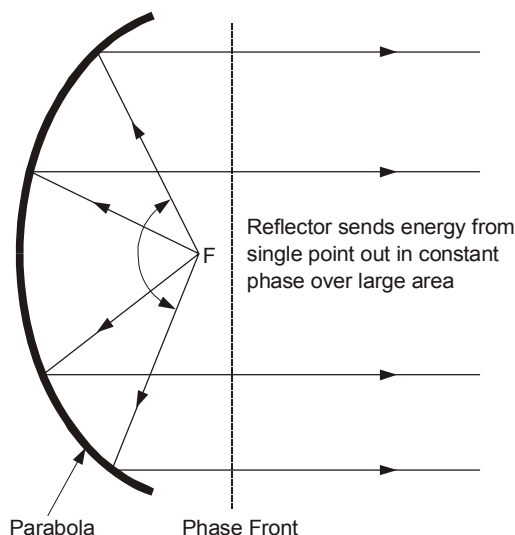
### Introduction

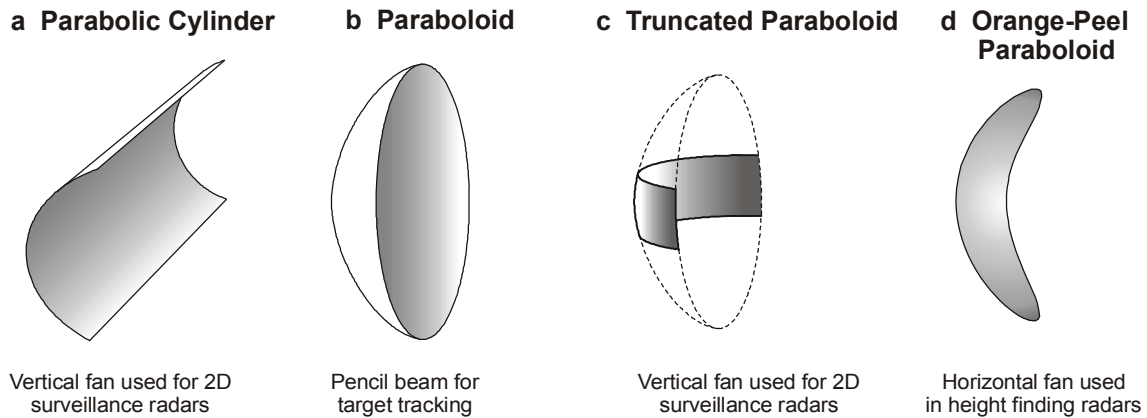
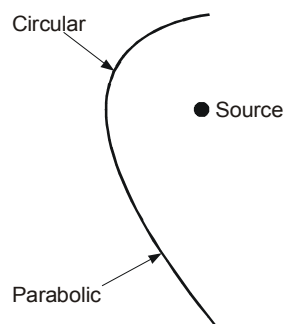
98. One of the principle advantages of operating in the microwave region of the spectrum is that it is possible to obtain a narrow beam width from a small aerial, allowing accurate bearings to be measured, and permitting rapid movement of the aerial, where necessary. The narrow beam may be formed by a reflector or a lens (in a similar manner to light), or by an array of radiating elements.

### Parabolic Reflectors and Feeds

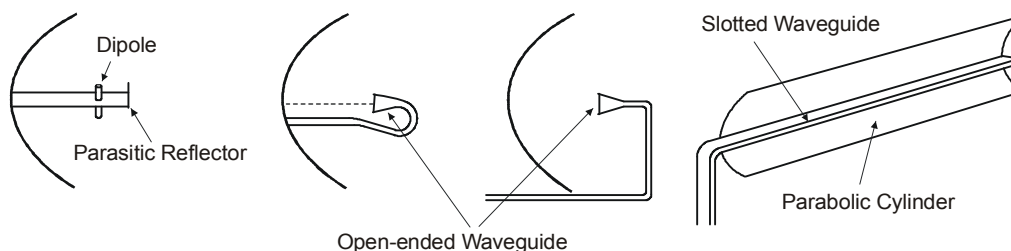
99. The parabolic reflector is perhaps the most familiar beam-forming device. If an aerial is placed at the focal point, energy radiated towards the reflector is reflected parallel to its axis (Fig 44). The larger the reflector, and the smaller the wavelength, the narrower the beam will be. Scanning or pointing of the beam is accomplished by mechanical movement of the aerial. Different horizontal and vertical beamwidths may be obtained by using different horizontal and vertical dimensions, and the reflector may be parabolic in both dimensions or only in one. Some typical reflector shapes, with their applications, are illustrated in Fig 45. The cosecant<sup>2</sup> shape, in which the top half of the dish has a circular cross-section, and the bottom half a parabolic cross section (Fig 46), is often used in airborne ground mapping radars. It is designed to ensure that all the ground to be mapped is illuminated with uniform energy and, to achieve this, more energy must be directed to long ranges than to short ranges. The required cosec<sup>2</sup> pattern may be also be obtained by placing a small prism of dielectric, or metal plates, in the aperture of a parabolic dish, or by using multiple stacked feeds.

14-20 Fig 44 Parabolic Reflector



**14-20 Fig 45 Some Typical Reflector Shapes and their Applications****14-20 Fig 46 Cosec<sup>2</sup> Dish**

**100. Feeds for Parabolic Aerials.** The feed for a parabolic aerial can only approximate to the theoretical point source. In practice, it may consist of a dipole placed ahead of a parasitic reflector element but, more commonly, an open-ended waveguide is used; some typical arrangements are shown in Fig 47. In the case of the parabolic cylinder, the axial feed usually consists of a length of slotted waveguide. The design of the feed must produce the desired power distribution across the aperture (typically a cosine pattern), and the waveguide plumbing must cause the minimum obscuration of the aperture.

**14-20 Fig 47 Feeds for Parabolic Aerials**

**101. Beamwidth and Sidelobes.** The beamwidth ( $\theta$ ) of a parabolic reflector may be approximated by:

$$\theta = \frac{70\lambda}{d}$$

where  $\lambda$  is the wavelength in use and  $d$  is the aperture diameter. Sidelobes of the order of 20 dB to 30 dB below the main beam are readily achieved and, with extreme care, it is possible to reduce sidelobes to 40 dB down.