# Electrical, Instruments & Radio

(According to the Syllabus Prescribed by Director General of Civil Aviation, Govt. of India)

# FIRST EDITION

# ELECTRICAL, INSTRUMENTS & RADIO


**Prepared by**

L.N.V.M. Society Group of Institutes

* School of Aeronautics
( Approved by Director General of Civil Aviation, Govt. of India)

* School of Engineering & Technology
( Approved by Director General of Civil Aviation, Govt. of India)

**Compiled by**
G.Mandal, A.K. Datta, Ashok Atri, M.P. Sharma,
Sandeep Kumar Verma, Shanida Mol S.



**Published By**

L.N.V.M. Society Group of Institutes
**H-974, Palam Extn., Part-1, Sec-7, Dwarka, New Delhi-45**

**Published By**
L.N.V.M. Society Group of Institutes,
Palam Extn., Part-1, Sec.-7,
Dwarka, New Delhi - 45

**First Edition 2007**



**Type Setting**
Anita Gupta

**Cover Designed by**
Abdul Aziz

**Printed at Graphic Syndicate, Naraina, New Delhi**

Dedicated To

Shri. Laxmi Narain Verma

[ Who Lived An Honest Life ]

# Preface

Avionics (Electrical, Instruments and Radio) is considered the life line of Aviation. Pilot remains in the command of aircraft with the help of the avionics, rather it may be said avionics guides the Pilot to control aircraft safely and to navigate to it's destination. With the steady growth of aircraft and their systems the design, constructional patterns and presentation of data have grown in a complex fashion in the avionics field.

There is, of course, no denying the above facts but every evolutionary process depends on well established principles.

Thus in preparing the material for this book, attention has been given to the fundamental principles of electricity and their application in aircraft, emphasis has been placed on the understanding of application of principles of basic physical laws associated with the Aircraft and Powerplant Instruments, their constructional details and function. Construction and functional details of radio system has been incorporated in a simple understandable language. The book is decorated with the functional diagram of the topics covered for easy comprehension.

This book is prepared by L.N.V.M. Society Group of Institutes with the unlenting efforts of it's vast experienced faculty and staff with the view to present the essence of the AVIONICS in a capsule form.

The subject details under single cover will open the door of knowledge of Electrical, Instrument and Radio systems to the aspirants of Aircraft Maintenance Engineer and will help them to get through their DGCA Licence Paper-II and better understanding of the fundamentals will help them in their professional life too.

Director Mr. C.C. Ashoka is a dynamic and inspiring force behind the publishing of this book.

My thanks are due to those who helped me to bring out this valuable edition.

I would very much appreciate criticism, suggestions and detection of errors from the readers which will be gratefully acknowledged.

**G. Mandal**

Senior Instructor

L.N.V.M. Society Group of Institutes

Dated : Dec. 2006

# SYLLABUS COVERED IN THIS BOOK ONLY FOR DGCA LICENCE PAPER - II EXAMINATION

**Knowledge of electrical terminology and components used in AC/DC circuitry, Ohm's law, Kirchhoff's law and their application. Principles of electromagnetic Induction and their application. Various methods of voltage regulation. Principle of operation of electrical test equipment.**

**Knowledge of Batteries and their maintenance.**

**Knowledge of principle of operation of aircraft and engine instruments.**

**Knowledge of various types of diodes/triodes/transistors and their functions.**

**Knowledge of conversion from decimal to binary system and vice versa. Symbols in logic gates.**

**Elementary knowledge of computer, its applications.**

**Identify the bands of frequency spectrum, their uses and propagation characteristics.**

# CONTENTS


## ELECTRICAL

| | | |
|---|---|---|
| 1. | ELECTRICAL AND MAGNETIC QUANTITIES, DEFINITIONS AND UNITS | 03 |
| 2. | OHM'S LAW | 09 |
| 3. | KIRCHHOFF'S LAW | 32 |
| 4. | ELECTROMAGNETIC INDUCTION | 38 |
| 5. | DIRECT CURRENT GENERATOR | 51 |
| 6. | ALTERNATING CURRENT GENERATORS | 72 |
| 7. | DC MOTORS | 78 |
| 8. | AC MOTORS | 84 |
| 9. | VOLTAGE REGULATION | 94 |
| 10. | AIRCRAFT BATTERIES | 101 |
| 11. | POWER CONVERSION EQUIPMENT | 118 |
| 12. | AIRCRAFT ELECTRICAL TEST EQUIPMENTS | 128 |
| 13. | ELECTRICAL DIAGRAM SYMBOLS | 151 |

## INSTRUMENT

| | | |
|---|---|---|
| 14. | AIRCRAFT INSTRUMENT PANELS AND RANGE MARKING | 155 |
| 15. | FLIGHT INSTRUMENTS-PITOT-STATIC SYSTEMS | 158 |
| 16. | FLIGHT INSTRUMENTS-GYROSCOPIC SYSTEMS | 167 |
| 17. | DIRECT-READING MAGNETIC COMPASSES | 174 |
| 18. | REMOTE READING COMPASSES | 176 |
| 19. | ENGINE INSTRUMENTS | 187 |
| 20. | AIRCRAFT RADIO NAVIGATION INSTRUMENTS | 208 |
| 21. | AUTO PILOT | 214 |
| 22. | ELECTRONIC (CRT) DISPLAYS | 219 |

## ELECTRONICS/RADIO


| | | |
|---|---|---|
| 23. | SEMICONDUCTOR DIODES | 229 |
| 24. | TRIODES | 245 |
| 25. | BIPOLAR JUNCTION TRANSISTOR | 248 |
| 26. | TRANSISTOR AMPLIFIERS | 262 |
| 27. | DECIMAL PREFIXES | 275 |
| 28. | DECIMAL TO BINARY CONVERSION AND VICE -VERSA | 276 |
| 29. | LOGIC GATES AND TRUTH TABLES | 287 |
| 30. | ELEMENTARY KNOWLEDGE OF COMPUTERS, ITS APPLICATIONS | 288 |
| 31. | SPECTRUM OF WAVES | 295 |
| 32. | COMMUNICATIONS SYSTEM | 310 |


■■■

# ELECTRICAL SYSTEM

# CHAPTER : 1
# ELECTRICAL AND MAGNETIC QUANTITIES, DEFINITIONS AND UNITS

| Quantity | Definition | Name of Unit | Unit symbol | Unit definition |
|---|---|---|---|---|
| Electric potential | That measured by the energy of a unit positive charge at a point, expressed relative to zero potential, or earth. | | | Difference of electric potential between two points of a conductor carrying constant current of 1 ampere, when the power dissipated between these points is equal to 1watt |
| Potential difference (p.d) | That between two points when main tained by an e.m.f., or by a current flowing through a resistance. | Volt | V | |
| Electromotive force (e.m.f.) | Difference of potential produced by sources of electrical energy which can be used to drive currents through external circuits. | | | |
| Current | The rate of flow of electric charge at a point in a circuit. | Ampere<br>Milliampere($A \times 10^{-3}$)<br>Microampere($A \times 10^{-6}$) | A<br>mA<br>$\mu$ A | The ampere is that constant current which , if maintained in two straight parallel conductors of infinite length, of negligible cross section, and placed 1 metre apart in vacuum, would produce between the conductors a force equal to 2 x $10^{-7}$ newton per metre of length. |
| Resistance | The tendency of a conductor to oppose the flow of current and to con vert electrical energy into heat. Its magnitude depends on such factors as: nature of conductor material, its physical state, dimensions, temperature and thermal properties; frequency of current and its magnitude. | Ohm<br>Mega ohm ($\Omega \times 10^6$) | $\Omega$<br>M$\Omega$ | The ohm is the electrical resistance between two points of a conductor when a constant p.d. of 1 volt, applied to these points, produces in the conductor a current of 1 ampere, the conductor not being the source of any e.m.f. |
| Power | The rate of doing work or transform-ing energy | Watt<br>Kilowatt ($W \times 10^3$) | W<br>kW | Is the power which in 1 second gives rise to energy of 1 joule. |
| Frequency | The number of cycles in unit time | Hertz | Hz | The definition of frequ-ency also applies with the unit of time being taken as 1 second. |
| Inductance | The property of an element or circuit which, when carrying a current, is characterized by the formation of a magnetic field and the storage of magnetic energy | Henry | H | The inductance of a closed circuit in which an e.m.f of 1 volt is produced when the current in the circuit varies at the rate of 1 ampere per second. |
| Capacitance | The property of a system of conduc-tors and insulators (a system known | Farad<br>Microfarad ($F \times 10^{-6)}$ | F<br>$\mu$F | The capacitance of capa-citor between the plates |

| | | | | |
|---|---|---|---|---|
| | as a capacitor) which allows the storage of an electric charge when a p.d exists between the conductors. In a capacitor, the conductors are known as electrodes or plates, and the insulator, which may be solid, liquid or gaseous is known as the dielectric. | Picofarad ($F \times 10^{-12}$) | pF | of which there appears a p.d of 1 volt when it is charged by a quantity of of electricity of 1 coulomb |
| Electric Charge | The quantity of electricity on an electrically charged body, or passing at a point in an electric circuit during a given time. | Coulomb | C | The quantity of electricity carried in 1 second by a current of 1 ampere |
| Energy | The capacity for doing work | Joule | J | The work done when the point of application of a force of 1 newton is displaced through a distance of 1 metre in the direction of the force. |
| Impedence | The extent to which the flow of alternating current at a given frequency is restricted and represented by the ratio of r.m.s. values of voltage and current. Combines resistance, capacitive and inductive reactance. | Ohm | Z | |
| Reactance | That part of the impedence which is due to inductance or capacitance, or both, and which stores energy rather than dissipates it. | Ohm | X | |
| Magnetic flux | A phenomenon produced in the medium surrounding electric currents or magnets. The amount of flux through any area is measured by the quantity of electricity caused by flow in a circuit of given resistance bounding the area when this circuit is removed from the magnetic field. | Weber (Volt-second) | Wb | The magnetic flux which, linking a circuit of 1 turn, would produce in it an e.m.f of 1 volt if it were reduced to zero at a uniform rate in 1 second. |
| Magnetic flux density (Magnetic induction) | The amount of magnetic flux per square metre, over a small area at a point in a magnetic field. The direction of the magnetic flux is at right angles to the area. | Tesla | T | Equal to 1 weber per square metre of circuit area. |
| Magnetic field strength (Magnetizing force) | The strength or force which produces or is associated with magnetic flux density. It is equal to the magnetomotive force per metre measured along the line of force. | Ampere per metre | A/m | |
| Magnetomotive force (m.m.f.) | The magnetic analogue of e.m.f. It represents the summated current or equivalent current, including any displacement current, which threads a closed line in a magnetic field and produces a magnetic flux along it. Can also be stated as the work done in moving a unit magnetic pole around a closed magnetic circuit. | Ampere-turns<br>Gilbert | | The product of current and the number of turns of a coil. |
| Reluctance | The ratio of magnetic force to magnetic flux. May be considered as the opposition to the flux established by the force. It is the reciprocal of permeance. | Ampere-turn/Weber (Gilbert/Maxwell) | | |

| | | | |
|---|---|---|---|
| Permeability ($\mu$) | The ratio of the magnetic flux density in a medium to the magnetizing force producing it. | Henry per metre (Weber per ampere-metre) | H/m (Wb/A-m) |
| Permeance | The capability of a magnetic circuit to produce a magnetic flux under the influence of an m.m.f., and which is represented as the quotient of a given magnetic flux in the magnetic circuit and the m.m.f. required to produce it. | | |

## ELECTRIC CIRCUITS AND NETWORK THEOREMS

There are certain theorems, which when applied to the electric networks, either simplify the network itself or render their analytical solution very easy. These theorems can also be applied to an a.c. system, with the only difference that impedences replace the ohmic resistance of d.c. system. Different electric circuits (according to their properties) are defined below:

### Node

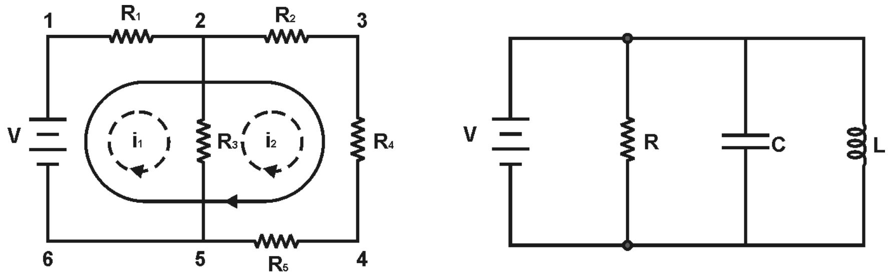
Node is a junction in a circuit where two or more circuit elements are connected together.

### Branch

Branch is that part of a network which lies between two junctions.

### Mesh

It is a loop that contains no other loop within it. For example, the circuit of Fig.1.1 (a) has seven branches, six nodes, three loops and two meshes whereas the circuit of Fig.1.1 (b) has four branches, two nodes, six loops and three meshes.




(a) (b)

*Fig.1.1*

### Loop

It is a close path in a circuit in which no element or node is encountered more than once.

## TYPES OF CIRCUIT CONNECTION

### Short and Open Circuits

When two points of circuit are connected together by a thick metallic wire (Fig. 1.2), they are said to be short-circuited. Since 'short' has practically zero resistance, it gives rise to two important facts:

I. No voltage can exist across it because V = IR = I X 0 = 0

II. Current through it (called short-circuit current) is very large (theoretically, infinity)

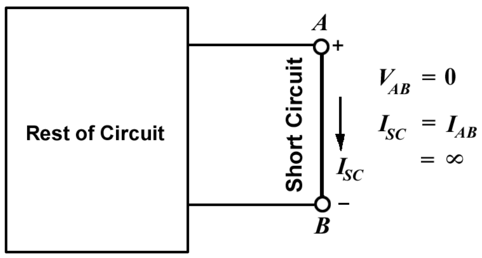



*Fig.1.2*

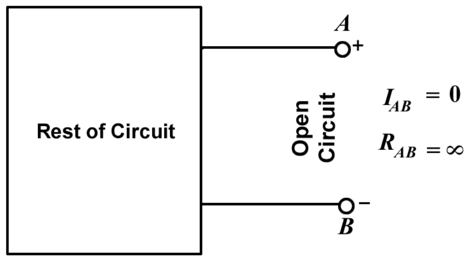



*Fig.1.3*

Two points are said to be open-circuited when there is no direct connection between them (Fig. 1.3). Obviously, an 'open' represents a break in the continuity of the circuit. Due to this break :

I. Resistance between the two points is infinite.

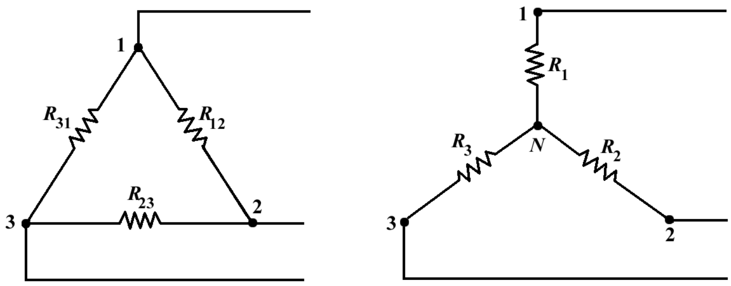
II. There is no flow of current between the two points.

**Delta/Star* Transformation**

In solving networks (having considerable number of branches) by the application of Kirchhoff's Law, one sometimes experiences great difficulty due to a large number of simultaneous equations that have to be solved. However, such complicated network can be simplified by successively replacing delta meshes by equivalent star system and *vice versa*.

Suppose we are given three resistances $R_{12}$, $R_{23}$ $R_{31}$ connected in delta fashion between terminals 1, 2 and 3 as in Fig. 1.4 (a). So far as the respective terminals are concerned, these three given resistances can be replaced by the three resistances $R_1$, $R_2$ $R_3$ connected in star as shown in Fig. 1.4 (b).

These two arrangements will be electrically equivalent if the resistance as measured between any pair of terminals is the same in both the arrangements. Let us find this condition.




*(a)* *(b)*

*Fig. 1.4*

First, take delta connection : Between terminal 1 and 2, there are two parallel paths; one having a resistance of $R_{12}$ and the other having a resistance of $(R_{23} + R_{31})$.

$\therefore$ Resistance between terminals 1 and 2 is $= \dfrac{R_{12} \times (R_{23} + R_{31})}{R_{12} + (R_{23} + R_{31})}$

Now, take star connection : The resistance between the same terminals 1 and 2 is $(R_1 + R_2)$.
As terminal resistances have to be the same

$$\therefore \quad R_1 + R_2 = \frac{R_{12} \times (R_{23} + R_{31})}{R_{12} + R_{23} + R_{31}} \quad \text{...(i)}$$

Similarly, for terminals 2 and 3 and terminals 3 and 1, we get

$$R_2 + R_3 = \frac{R_{23} \times (R_{31} + R_{12})}{R_{12} + R_{23} + R_{31}} \quad \text{...(ii)}$$

and

$$R_3 + R_1 = \frac{R_{31} \times (R_{12} + R_{23})}{R_{12} + R_{23} + R_{31}} \quad \text{...(iii)}$$

Now, subtracting (ii) from (i) and adding the result to (iii), we get

$$R_1 = \frac{R_{12}\, R_{31}}{R_{12} + R_{23} + R_{31}}; R_2 = \frac{R_{23}\, R_{12}}{R_{12} + R_{23} + R_{31}} \text{ and } R_3 = \frac{R_{31}\, R_{23}}{R_{12} + R_{23} + R_{31}}$$

**How to Remember?**

It is seen from above that each numerator is the product of the two sides of the delta which meet at the point in star. Hence, it should be remembered that : *resistance of each arm of the star is given by the product of the resistances of the two delta sides that meet at its end divided by the sum of the three delta resistances.*

**Star/Delta Transformation**

This transformation can be easily done by using equations (i), (ii) and (iii) given above. Multiplying (i) and (ii), (ii) and (iii), (iii) and (i) and adding them together and then simplifying them, we get

$$R_{12} = \frac{R_1R_2 + R_2R_3 + R_3R_1}{R_3} = R_1 + R_2 + \frac{R_1 R_2}{R_3}$$

$$R_{23} = \frac{R_1R_2 + R_2R_3 + R_3R_1}{R_1} = R_2 + R_3 + \frac{R_2 R_3}{R_1}$$

$$R_{31} = \frac{R_1R_2 + R_2R_3 + R_3R_1}{R_2} = R_1 + R_3 + \frac{R_3 R_1}{R_2}$$

**How to Remember?**
The equivalent delta resistance between any two terminals is given by the sum of star resistances between those terminals plus the product of these two star resistances divided by the third star resistance.

**Ideal Constant-Voltage Source**
It is that voltage source (or generator) whose output voltage remains absolutely constant whatever be the change in load current. Such a voltage source must possess *zero internal resistance so that internal voltage drop in the source is zero.* In that case, output voltage provided by the source would remain constant *irrespective of the amount of current drawn from it.* In practice, none such ideal constant-voltage source can be obtained. However, smaller the internal resistance r of a voltage source, closer it comes to the ideal sources described above.

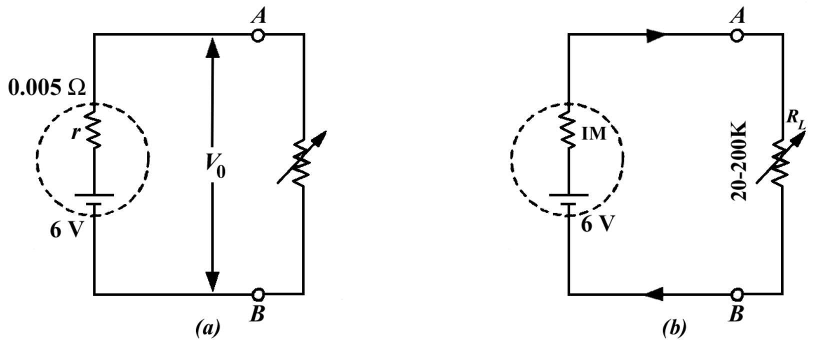



*Fig.1.5*

Suppose, a 6-V battery has an internal resistance of 0.005 Ω [Fig. 1.5 (a)]. When it supplies no current i.e. it is on no-load, $V_o$ = 6 V i.e. output voltage provided by it at its output terminals A and B is 6 V. If load current increases to 100 A, internal drop = 100 X 0.005 = 0.5 V. Hence, $V_o$ = 6 - 0.5 = 5.5 V.

Obviously an output voltage of 5.5 - 6 V can be considered constant as compared to wide variations in load current from 0 A to 100 A.

**Ideal Constant-Current Source**
It is that voltage source whose internal resistance is infinity. In practice, it is approached by a source which possess very high resistance as compared to that of the external load resistance. As shown in Fig. 1.5 (b), let the 6-V battery or voltage source have an internal resistance of 1 MΩ and let the load resistance vary from 20 K to 200 K. The current supplied by the source varies from 6.1/1.02 = 5.9 μA to 6/1.2 = 5 μA. Hence, the source can be considered, for all practical purposes, to be a constant-current source.

**Source Conversion**
A given voltage source with a series resistance can be converted into (or replaced by) an equivalent current source with a parallel resistance. Conversely, a current source with a parallel resistance can be converted into a voltage source with a series resistance. Suppose, we want to convert the voltage source of Fig. 1.6 (a) into an equivalent current source. First, we will find the value of current supplied by the source when a 'short' is put across terminals A and B as shown in Fig. 1.6 (b). This current is I = V/R.

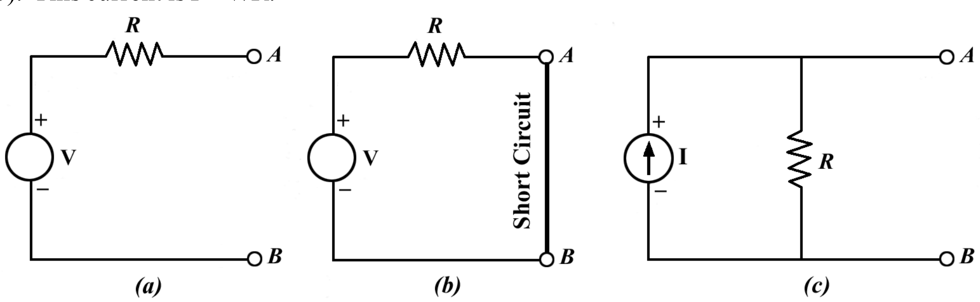



*Fig.1.6*

A current source supplying this current *I* and having the same resistance R connected in parallel with it represents the equivalent source. It is shown in Fig. 1.6 (c). Similarly, a current source of I and a parallel resistance R can be converted into a voltage V = IR and a resistance R in series with it. It should be kept in mind that a voltage source-series resistance combination is equivalent to (or replaceable by) a current source-parallel resistance combination if, and only if their

1. Respective open-circuit voltages are equal, and
2. Respective short-circuit currents are equal.

For example, in Fig.1.6 (a), voltage across terminals A and B when they are open (i.e. open-circuit voltage $V_{oc}$) is V itself because there is no drop across R. Short-circuit current across AB = 1 = V/R.

Now, take the circuit of Fig. 1.6 (c). The open-circuit voltage AB = drop across R = IR = V. If a short is placed across AB, whole of I passes through it because R is completely shorted out.

■■■

# CHAPTER : 2
# OHM'S LAW

This law is fundamental to all direct current circuits, and can in a modified form, also be applied to alternating current circuits.

The law may be stated as follows: When current flows in a conductor, the difference in potential between the ends of the conductor, divided by the current flowing, is a constant provided there is no change in the physical condition of the conductor.

The constant is called the resistance (R) of the conductor, and is measured in ohms (Ω). In symbols,

$$R = \frac{V}{I} \qquad (1)$$

Where,

V = Potential difference in volts.

I = Current in amperes.

Calculations involving most conductors, either single or in a variety of combinations are easily solved by this law, for if any two of the three quantities (V, I and R) are known, the third can always be found by simple transposition. Thus, from (1)

$$V = IR \text{ volts} \qquad (2)$$

$$I = \frac{V}{R} \text{ amperes} \qquad (3)$$

**POWER**

Since some of the power delivered to a circuit are the same as those used in Ohm's law, it is possible to substitute Ohm's law values for equivalents in the fundamental formula for power (P) which is:

$$P = V \times I \text{ watts}$$

Thus, if V/R is substituted for I in the power formula, it becomes

$$P = V \times \frac{V}{R} \quad \text{or} \quad P = \frac{V^2}{R}$$

Similarly, if IR is substituted for V in the power formula, then

$$P = I \times I \times R \qquad \text{or} \qquad P = I^2 R$$

By transposing the formula $P = I^2 R$ to solve for the current I, we obtain

$$I^2 = \frac{P}{R}$$

from which

$$I = \sqrt{\frac{P}{R}}$$

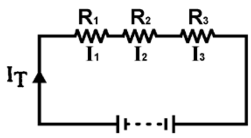
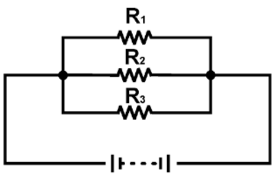
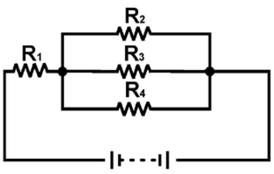
Other transpositions of the foregoing formula are as follows:

$$I = \frac{P}{V} \qquad V = \sqrt{P.R}$$

$$R = \frac{P}{I^2}$$

$$V = \frac{P}{I}$$

**APPLICATION OF OHM'S LAW TO SERIES AND PARALLEL CIRCUITS (RESISTANCES)**

| Circuit | Total resistance | Total voltage | Total current |
|---|---|---|---|
| <br>Series | $R_T = R_1 + R_2 + R_3 + \ldots$ ohms<br>or $R_T = \frac{V_T}{I}$ ohms<br>If the resistances are of equal value R then :<br>$R_T = nR$ ohms<br>where n = number of resistors | $V_T = (I_1R_1) + (I_2R_2) + (I_3R_3) + \ldots$ volts<br>or $V_T = IR_T$ volts | $I_T = I_1 = I_2 = I_3 = \ldots$ amps<br>or $I_T = \frac{V_T}{R_T}$ |
| <br>Parallel | $\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots$<br>or $R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots}$<br>or $R_T = \frac{V_T}{I_T}$<br>If the resistances are of equal value R, then:<br>$R_T = \frac{R}{n}$<br>When only two resistances in parallel, the total resistance is :<br>$R_T = \frac{R_1 \times R_2}{R_1 + R_2}$ | $V_T = V_1 = V_2 = V_3 = \ldots$ volts | $I_T = I_1 + I_2 + I_3 + \ldots$ amps |
| <br>Series-parallel | $R_T$, $V_T$ and $I_T$ are found by first reducing the parallel circuit to a single resistance, and then solving the whole as a simple series circuit. | | |

# NUMERICALS

1. If a resistor is to dissipate energy at the rate of 250 W, find its resistance for a terminal voltage of 100 V.

**Sol.:** Here; W = 250 watt, V = 100 volt; R = ?

$W = V^2/R$ or $R = V^2/W = 100^2/250 = 40\,\Omega$

2. In the bridge circuit of Fig, 2.1 calculate the reading of a voltmeter connected across (i) AB (ii) BC (iii) AD (iv) DC and (v) BD.

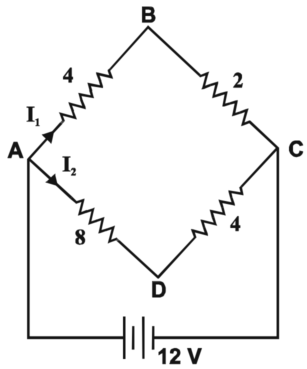



*Fig.2.1*

**Sol.:** The two branches ABC and ADC consist of two resistors connected in series. These two branches are connected in parallel across the 12 V battery. Taking branch ABC and applying Ohm's law, we have

$I_1 = 12/6 = 2A$

i) $V_{AB} = I_1 \times R_{AB} = 2 \times 4 = 8V$

ii) $V_{BC} = I_1 \times R_{BC} = 2 \times 2 = 4V$

$I_2 = 12/12 = 1A$

iii) $V_{AD} = I_2 \times R_{AD} = 1 \times 8 = 8V$
iv) $V_{DC} = I_2 \times R_{DC} = 1 \times 4 = 4V$
v) Since $V_{AB} = V_{AD}$, the points B and D are at the same potential .
Hence, a voltmeter connected across them will read zero.

3. An incandescent projector lamp has the rated voltage of 60 volts and hot resistance of 20Ω. Find the series resistance required to operate the lamp from a 75 volt supply.

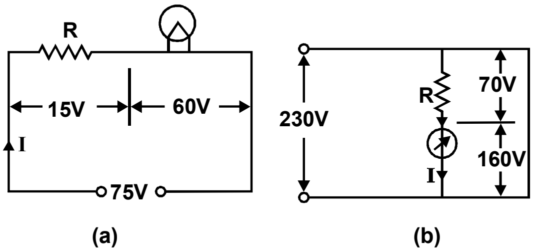



*Fig.2.2*

**Sol.:** Rated value of lamp current = 60/20 = 3A
Let R be the required series resistance [Fig.2.2 (a)].
Then, the excess voltage of (75-60) = 15 V is to be dropped on R.
R = 15/3 = 5Ω.

4. A voltmeter has a resistance of 20,000 Ω. When connected in series with an external resistance across a 230 V supply, the instrument reads 160 V. What is the value of external resistance ?
**Sol.:** The circuit is shown in Fig.2.2 (b). The voltage drop across external resistance R = 230 - 160 = 70 V
Circuit current I = 160/20,000 = 1/125 A
Now IR = 70
1/125 x R = 70 or R = 8,750 Ω

5. The resistance of two wires is 25 Ω when connected in series and 6 Ω when joined in parallel. Calculate the resistance of each wire.
**Sol.:** Let the two unknown resistances be $R_1$ and $R_2$ .
Then, when in series

$$R_1 + R_2 = 25 \quad \text{(i)}$$

When joined in parallel

$$(1/R_1) + (1/R_2) = 1/6 \quad \text{or} \quad 6 = [R_1R_2 / (R_1 + R_2)] \quad \text{(ii)}$$

Putting the value of $R_2$ from Eq. (i) in Eq. (ii), we have

$$6 = [R_1 (25 - R_1) / 25]$$

or $R_1^2 - 25 R_1 + 150 = 0$
or $(R_1 - 15)(R_1 - 10) = 0$
$R_1 = 10\,\Omega$, so $R_2 = 15\,\Omega$ or $R_1 = 15\,\Omega$ so $R_2 = 10\,\Omega$
Hence, the two wires have resistances of 10 Ω and 15 Ω.

6. The equivalent resistance of four resistors joined in parallel is 20 Ω. The currents flowing through them are 0.6, 0.3, 0.2 and 0.1 A. Find the value of each resistor.

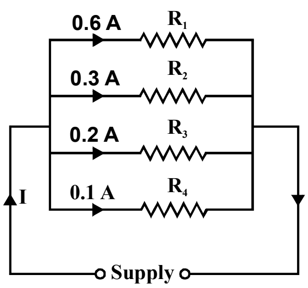



*Fig.2.3*

**Solution :** Total current in the circuit is the sum of the four branch currents. Its value is = 0.6 + 0.3 + 0.2 + 0.1 = 1.2 A
The common voltage across parallel resistors = 20 x 1.2 = 24 V (Fig. 2.3)
∴ $R_1 = 24/0.6 = 40\,\Omega$
$R_2 = 24/0.3 = 80\,\Omega$
$R_3 = 24/0.2 = 120\,\Omega$
$R_4 = 24/0.1 = 240\,\Omega$

7. Calculate the effective resistance of the following combination of resistances and the voltage drop across each resistance when a p.d. of 60 V is applied between points A and B. Figures represent resistances in ohms.

**Solution :** Resistance between A and C (Fig. 2.4) is

$= 6 \parallel 3 = 6 x 3/6+3 = 2\Omega$

Resistance of branch ACD = 18 + 2 = 20 Ω

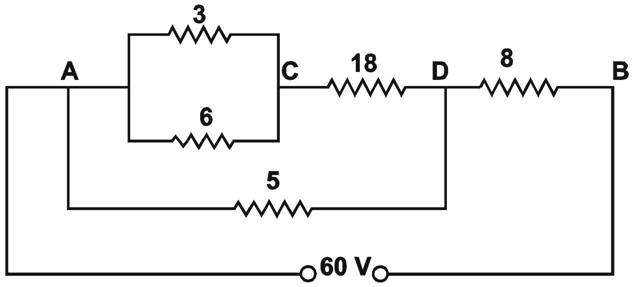



*Fig.2.4*

Now, there are two parallel paths between points A and D of resistances 20 Ω and 5 Ω. Hence, resistance between A and D

$= 20 \parallel 5 = 20 x 5/20+5 = 4\,\Omega$

resistance between A and B = 4 + 8 = 12 Ω

Total circuit current = 60/12 = 5 A

Current through 5 Ω resistor = 5 x 20/25 = 4 A

Current in branch ACD = 5 x 5/25 = 1A

P.D. across 3 Ω and 6 Ω resistor = 1 x 2 = 2 V

P.D. across 18 Ω resistor = 18 x 1 18 V

P.D. across 5 Ω resistor = 5 x 4 = 20 V

P.D. across 8 Ω resistor = 8 x 5 = 40 V

8. In the circuit shown in Fig.2.5, find the voltage across and the current in each element.

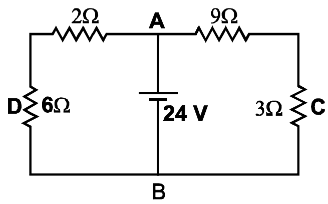



*Fig.2.5*

**Solution**. As seen, there are two parallel paths between points A and B ; one path is ACB and has a total resistance of (9+3) = 12 Ω. The other is ADB and has a total resistance of (2+6) = 8 Ω . Both paths have a p.d. of 24 V applied across them.

Current through path ACB is

= 24/12 = 2 A

Drop across 9 Ω = 2 x 9 = 18V

Drop across 3 Ω = 2 × 3 = 6 V

Current through path ADB is = 24/8 = 3A

Drop across 2Ω = 3 x 2 = 6 V; Drop across 6 Ω = 3 x 6 = 18 V

9. A 100 W, 250 V bulb is put in series with a 40 W, 250 V bulb across 500 V supply (i) what will be the current drawn (ii) what will be the power consumed by each bulb and (iii) will such a combination work ?

**Solution.** If $R_1$ and $R_2$ are the resistances of the two bulbs, then using the relation $W = V^2/R$, we have

$R_1 = V^2/W_1 = 250^2/100 = 625\,\Omega$, $R_2 = 250^2/40 = 1562.5\,\Omega$

(i) $I = 500/(625 + 1562.5) = 0.228\,A$

Incidentally, this current is less than the normal current of 100/250 = 0.4 A for 100 W bulb but more than the normal current of 40 / 250 = 0.16 A for the 40 W bulb. Hence, 100 W bulb will under glow whereas 40 W bulb will over-glow as proved below.

(ii) Power consumed by 100 W bulb is

$= I^2 R = (0.228)^2 x 625 = 32.5$ W

Hence, this bulb will operate at nearly one-third its normal glow.

Power consumed by 4 - W bulb is

$= (0.228)^2$ x 1562.5 = 81 W

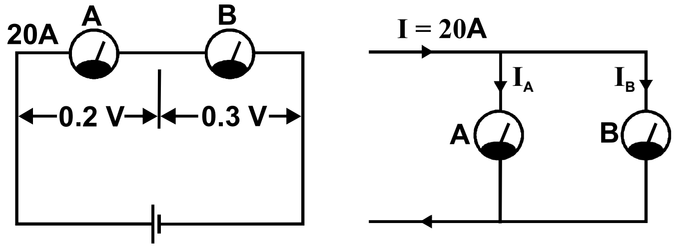
Obviously, this bulb will operate with twice the normal glow. Hence, it would burn out very quickly.

(iii) This combination will not operate properly for the reasons given in (ii) above.

10. A current of 20 A flows through two ammeters A and B joined in series. Across A, the potential difference is 0.2 V and across B it is 0.3 V. Find how the same current will divide between A and B when they are joined in parallel.
**Solution** The two ammeters are connected in series in Fig. 2.6 (a) $R_A = 0.2/20 = 0.01\ \Omega$ ; $R_B = 0.3/20 = 0.015\ \Omega$
The same two ammeters are connected in parallel in Fig.2.6 (b)

$$I_A = I \times \frac{R_B}{R_A + R_B} = 20 \times \frac{0.015}{0.025} = 12\text{ A}$$

$$I_B = I \times \frac{R_A}{R_A + R_B} = 20 \times \frac{0.01}{0.025} = 8\text{ A}$$




(a) (b)

*Fig.2.6*

11 A resistance of 10 Ω is connected in series with two resistances each of 15 Ω arranged in parallel. What resistance must be shunted across this parallel combination so that the total current taken shall be 1.5 A with 20 V applied ?
**Solution.** The original circuit is shown in Fig 2.7 Let R be the required resistance as shown in Fig 2.8. If $R_{eq}$ is the equivalent resistance of the parallel group, then

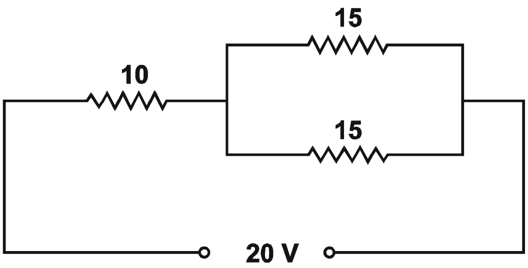



*Fig.2.7*

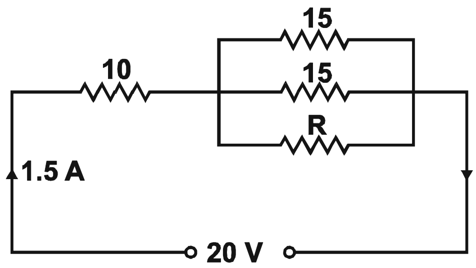
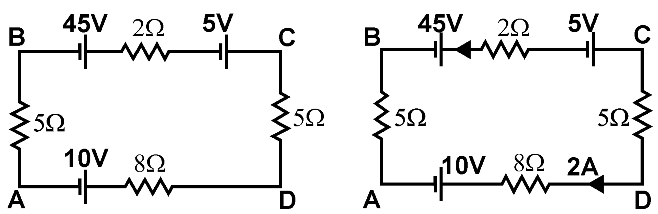



*Fig.2.8*

$$1.5 = \frac{20}{10 + R_{eq}} \qquad R_{eq} = 10/3\ \Omega$$

$$\frac{3}{10} = \frac{1}{15} + \frac{1}{15} + \frac{1}{R} \qquad R = 6\ \Omega$$

12 In the circuit shown in Fig.2.9 (a), determine the voltage rise from A to C and the power absorbed by the portion AD.
**Solution.** It should be noted that 45 V and 5 V batteries are connected in additive series whereas 10 V battery is connected in opposition to them i.e. in subtractive series.




(a) (b)

*Fig.2.9*

Hence net driving voltage around the circuit = (45 + 5) - 10 = 40 V. Total resistance = (2 + 5 + 8 + 5) = 20 Ω . Circuit current

= 40/20 = 2A. As shown, it flows clockwise round the circuit. Since current flows from C to D to A, it is obvious that C is at a higher electrical potential than A.

As we go from A to C via point D, we meet the following voltages:

i) Since we go from -ve to +ve terminal of the 10 V battery, there is an increase of 10 V.

ii) There is a voltage rise of $8 \times 2 = 16$ V over the 8 Ω resistance. This represents an increase or rise in voltage because we are going upstream i.e. opposite to the direction of flow of current.

iii) Similarly, there is a rise in voltage of 2 x 5 = 10 V as we go from D to C.

Total increase or rise in voltage is

$= 10 + 16 + 10 = 36$ V

**Note :** If we go from A to C via point B, the change in voltage is

$= -(5 \times 2) + 45 - (2 \times 2) + 5 = +36$ V

The positive sign indicates that it is a rise in voltage.

## RESISTANCE

It may be defined as the property of a substance due to which it opposes (or restricts) the flow of electricity (i.e., electrons) through it.

Metals (as a class), acids and salts solutions are good conductors of electricity. Amongst pure metals, silver, copper and aluminium are very good conductors in the given order. This, as discussed earlier, is due to the presence of a large number of free or loosely-attached electrons in their atoms. These vagrant electrons assume a directed motion on the application of an electric potential difference. These electrons while flowing pass through the molecules or the atoms of the conductor, collide with other atoms and electrons, thereby producing heat.

Those substances which offer relatively greater difficulty or hindrance to the passage of these electrons are said to be relatively poor conductors of electricity like bakelite, mica, glass, rubber, p.v.c. (polyvinyl chloride) and dry wood etc. Amongst good insulators can be included fibrous substances such as paper and cotton when dry, mineral oils free from acids and water, ceramics like hard porcelain and asbestos and many other plastics besides p.v.c. It is helpful to remember that electric friction is similar to friction in Mechanics

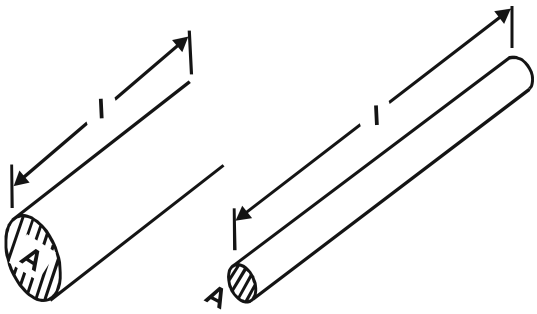
## THE UNIT OF RESISTANCE

The practical unit of resistance is ohm. A conductor is said to have a resistance of one ohm if it permits one ampere current to flow through it when one volt is impressed across its terminals.

For insulators whose resistances are very high, a much bigger unit is used i.e. Megaohm = $10^6$ ohm (the prefix 'Mega' or Mego meaning a million) or kilo ohm = $10^3$ ohm (kilo means thousand). In the case of very small resistances, smaller units like milli-ohm = $10^{-3}$ ohm or micro ohm = $10^{-6}$ ohm are used. The symbol for ohm is $\Omega$ .

**Table 2.1 Multiples and sub-multiples of Ohm**

| Prefix | Its meaning | Abbreviation | Equal to |
|---|---|---|---|
| Mega- | One million | MΩ | $10^6\Omega$ |
| Kilo- | One thousand | KΩ | $10^3\Omega$ |
| Centi- | One hundredth | - | $10^{-2}\Omega$ |
| Milli- | One thousandth | mΩ | $10^{-3}\Omega$ |
| Micro- | One millionth | μΩ | 10-6 Ω |




*Smaller - l* *Larger - l*
*Larger - A* *Smaller - A*
*Low - R* *Greater - R*

*Fig.2.10*

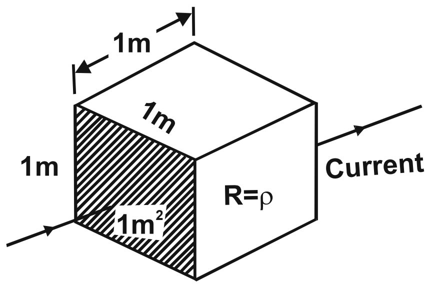



*Fig.2.11*

**LAWS OF RESISTANCE**

The resistance R offered by a conductor depends on the following factors :

(i) It varies directly as its length l.
(ii) It varies inversely as the cross-section A of the conductor.
(iii) It depends on the nature of the material.
(iv) It also depends on the temperature of the conductor.

Neglecting the last factor for the time being, we can say that

$$R \alpha \frac{\ell}{A} \text{ or } R = \rho \frac{\ell}{A} \qquad \text{(i)}$$

Where $\rho$ is a constant depending on the nature of the material of the conductor and is known as its specific resistance or resistivity.

If in eq. (i), we put

l = 1 metre and A = 1 metre$^2$, then $R = \rho$ (Fig. 2.11)

Hence, specific resistance of a material may be defined as the resistance between the opposite faces of a metre cube of that material.

**UNITS OF RESISTIVITY**

From Eq. (i), we have $\rho = \frac{AR}{\ell}$

In the S.I. system of units,

$$\rho = \frac{\text{A metre}^2 \times \text{R ohm}}{\ell \text{ metre}} = \frac{AR}{\ell} \text{ohm} - \text{metre}$$

Hence, the unit of resistivity is ohm-metre $(\Omega - \text{m})$.

It may, however, be noted that resistivity is sometimes expressed as so many ohm per m$^3$.
Although, it is incorrect to say so but it means the same thing as ohm-metre.

If l is in centimetres and A is cm$^2$, then $\rho$ is in ohm-centimetre $(\Omega - \text{cm})$.

**CONDUCTANCE AND CONDUCTIVITY**

Conductance (G) is reciprocal of resistance. Whereas resistance of a conductor measures the opposition which it offers to the flow of current, the conductance measures the inducement which it offers to its flow.

From Eq. (i), $R = \rho \frac{\ell}{A} \text{ or } G = \frac{1}{\rho} . \frac{A}{\ell} = \frac{\sigma A}{\ell}$

Where $\sigma$ is called the conductivity or specific conductance of a conductor. The unit of conductance is siemens (S). Earlier, this unit was called mho.

It is seen from the above equation that the conductivity of a material is given by

$$\sigma = G \frac{\ell}{A} = \frac{\text{G siemens } \times \ell \text{ metre}}{\text{A metre}^2} = G \frac{\ell}{A} \text{siemens / metre}$$

Hence, the unit of conductivity is siemens/metre (S/m).

**EFFECT OF TEMPERATURE ON RESISTANCE**

The effect of rise in temperature is :

(i) to increase the resistance of pure metals. The increase is large and fairly regular for normal ranges of temperature. The temperature/resistance graph is a straight line (Fig. 2.12). As would be presently clarified, metals have a positive temperature co-efficient of resistance.
(ii) to increase the resistance of alloys, though in their case, the increase is relatively small and irregular. For some high-resistance alloys like Eureka (60% Cu and 40% Ni) and manganin, the increase in resistance is (or can be made) negligible over a considerable range of temperature.
(iii) To decrease the resistance of electrolytes, insulators (such as paper, rubber, glass, mica etc.) and partial conductors such as carbon. Hence, insulators are said to possess a negative temperature-coefficient of resistance.

**TEMPERATURE COEFFICIENT OF RESISTANCE**

Let a metallic conductor having a resistance of $R_0$ at 0° be heated to t°C and let its resistance at this temperature be $R_t$. Then, considering normal ranges of temperature, it is found that the increase in resistance $\Delta R = R_t - R_0$ depends

(i) directly on its initial resistance
(ii) directly on the rise in temperature
(iii) on the nature of the material of the conductor.

or $R_t - R_0 \propto R_0 \times t$ or $R_t - R_0 = \alpha R_0 t$ (i)

where $\alpha$ (alpha) is a constant and is known as the temperature coefficient of resistance of the conductor.

Rearranging Eq. (i), we get $\alpha = \frac{R_t - R_0}{R_0 \times t} = \frac{\Delta R}{R_0 \times t}$

If $R_0 = 1\Omega, t = 1°C$, then $\alpha = \Delta R = R_t - R_0$

Hence, the temperature-coefficient of a material may be defined as :
the increase in resistance per ohm original resistance per °C rise in temperature.

From eq. (i), we find that $R_t = R_0(1 + \alpha t)$.

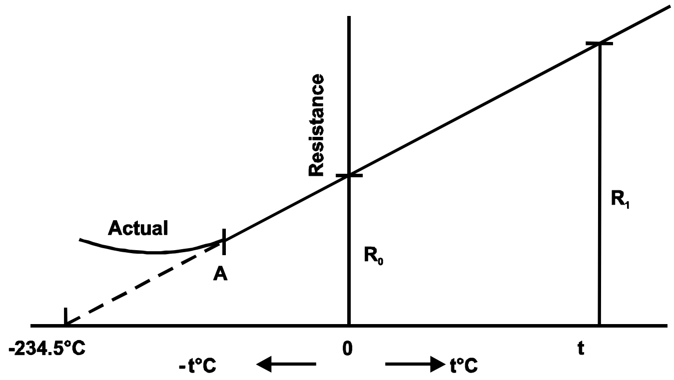



*Fig.2.12*

**4. TYPES OF RESISTORS**

**a. Carbon Composition**

It is a combination of carbon particles and a binding resin with different proportions for providing desired resistance. Attached to the ends of the resistive element are metal caps which have axial leads of tinned copper wire for soldering the resistor into a circuit. The resistor is enclosed in a plastic case to prevent the entry of moisture and other harmful elements from outside. Billions of carbon composition resistors are used in the electronic industry every year. They are available in power ratings of 1/8, 1/4, 1/2, 1 and 2 W, in voltage ratings of 250, 350 and 500 V. They have low failure rates when properly used.

Such resistors have a tendency to produce electric noise due to the current passing from one carbon particle to another. This noise appears in the form of a hiss in a loudspeaker connected to a hi-fi system and can overcome very weak signals. That is why carbon composition resistors are used where performance requirements are not demanding and where low cost is the main consideration. Hence, they are extensively used in entertainment electronics although better resistors are used in critical circuits.

**b. Deposited Carbon**

Deposited carbon resistors consist of ceramic rods which have a carbon film deposited on them. They are made by placing a ceramic rod in a methane-filled flask and heating it until, by a gascracking process, a carbon film is deposited on them. A helix-grinding process forms the resistive path. As compared to carbon composition resistors, these resistors offer a major improvement in lower current noise and in closer tolerance. These resistors are being replaced by metal film and metal glaze resistors.

**c. High-Voltage Ink Film**

These resistors consist of a ceramic base on which a special resistive ink is laid down in a helical band. These resistors are capable of withstanding high voltages and find extensive use in cathode ray circuits, in radar and in medical electronics. Their resistance ranges from 1 K$\Omega$ to 100,000 M$\Omega$ with voltage range upto 1000 kV.

**d. Metal Film**

Metal film resistors are made by depositing vaporized metal in vacuum on ceramic-core rod. The resistive path is helix-ground as in the case of deposited carbon resistors. Metal film resistors have excellent tolerance and temperature coefficient and are extremely reliable. Hence, they are very suitable for numerous high grade applications as in low-level stages of certain instruments although they are much more costlier.

**e. Metal Glaze**

A metal glaze resistor consists of a metal glass mixture which is applied as a thick film to a ceramic substrate and then fired to form a film. The value of resistance depends on the amount of metal in the mixture. With helix-grinding, the resistance can be made to vary from 1 $\Omega$ to many mega ohms.

Another category of metal glaze resistors consists of a tinned oxide film on a glass substrate.

**f. Wire-Wound**

Wire-wound resistors are different from all other types in the sense that no film or resistive coating is used in their construction. They consist of ceramic-core wound with a drawn wire having accurately-controlled characteristics. Different wire alloys are used for providing different resistance ranges. These resistors have highest stability and highest power rating.

Because of their bulk, high-power ratings and high cost, they are not suitable for low-cost or high-density, limited-space applications. The completed wire-wound resistor is coated with an insulating material such as baked enamel.

**g. Cermet (Ceramic Metal)**

The cermet resistors are made by firing certain metals blended with ceramics on a ceramic substrate. The value of resistance depends on the type of mix and its thickness. These resistors have very accurate resistance values and show high stability even under extreme temperatures. Usually, they are produced as small rectangles having leads for being attached to printed circuit boards (PCB).

**NONLINEAR RESISTORS**

Those elements whose V-I curves are not straight lines are called nonlinear elements because their resistances are nonlinear resistances.

Examples of nonlinear elements are filaments of incandescent lamps, diodes, thermistors and varistors. A varistor is a special resistor made of carborundum crystals held together by a binder. Fig. 2.13 (a) shows how current through a varistor increases rapidly when the applied voltage increases beyond a certain amount (nearly 100 V in the present case).

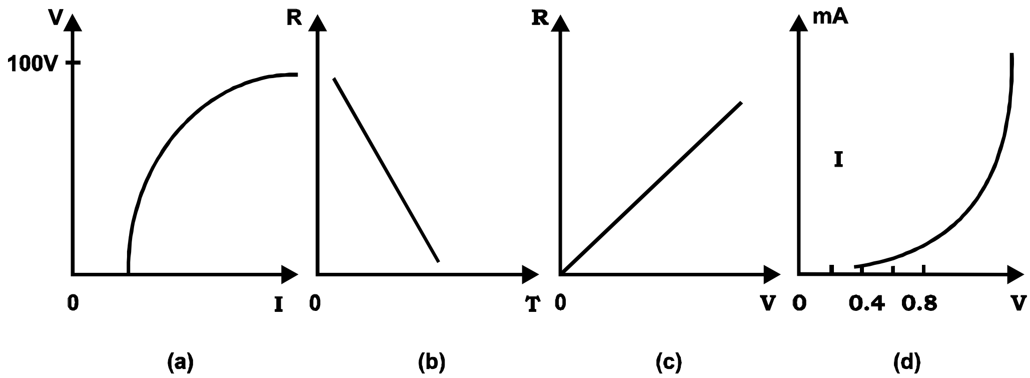



*Fig.2.13*

There is corresponding rapid decrease in resistance when the current increases. Hence, varistors are generally used to provide over-voltage protection in certain circuits.

A thermistor is made of metallic oxides in a suitable binder and has a large negative coefficient of resistance i.e. its resistance decreases with increase in temperature as shown in Fig. 2.13 (b). Fig. 2.13 (c) shows how the resistance of an incandescent lamp increases with voltage whereas Fig. 2.13 (d) shows the V-I characteristics of a typical silicon diode.

**VARISTOR (NONLINEAR RESISTOR)**

It is a voltage-dependent metal-oxide material whose resistance decreases sharply with increasing voltage. The relationship between the current flowing through a varistor and the voltage applied across it is given by the relation : $i = ke^{\eta}$ where i = instantaneous current, e is the instantaneous voltage and $\eta$ is a constant whose value depends on the metal oxides used. The value of $\eta$ for silicon-carbide-based varistor lies between 2 and 6 whereas zinc-oxide-based varistors have a value ranging from 25 to 50.

The zinc-oxide-based varistors are primarily used for protecting solid-state power supplies from low and medium surge voltage in the supply line. Silicon-carbide varistors provide protection against high-voltage surges caused by lighting and by the discharge of electromagnetic energy stored in the magnetic field of large coils.

## SHORT AND OPEN CIRCUITS

When two points of circuit are connected together by a thick metallic wire (Fig. 2.14), they are said to be short-circuited. Since 'short' has practically zero resistance, it gives rise to two important facts :

(i) no voltage can exist across it because $V = IR = I \times 0 = 0$.

(ii) current through it (called short-circuit current) is very large (theoretically, infinity).

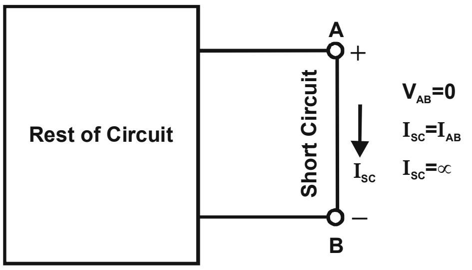



*Fig.2.14*

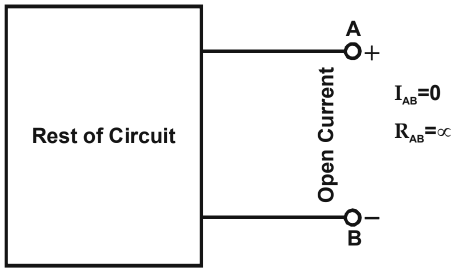



*Fig.2.15*

Two points are said to be open-circuited when there is no direct connection between them (Fig. 2.15). Obviously, an 'open' represents a break in the continuity of the circuit. Due to this break :

(i) resistance between the two points is infinite.

(ii) there is no flow of current between the two points.

## 'SHORT' IN A SERIES CIRCUIT

Since a dead (or solid) short has almost zero resistance, it causes the problem of excessive current which, in turn, causes power dissipation to increase many times and circuit components to burn out.

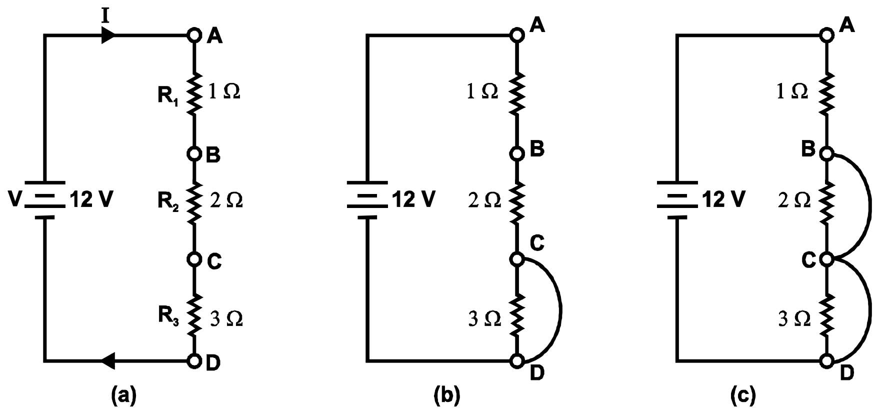



*Fig.2.16*

In Fig. 2.16 (a) is shown a normal series circuit where

$$V = 12\text{ V}, R = R_1 + R_2 + R_3 = 6\Omega$$

$$I = V/R = 12/6 = 2\text{A}, P = I^2R = 2^2 \times 6 = 24\text{ W}$$

In Fig. 2.16 (b), 3-$\Omega$ resistor has been shorted out by a resistance less copper wire so that $R_{CD} = 0$.

Now, total circuit resistance $R = 1 + 2 + 0 = 3\Omega$. Hence, I = 12/3 =4 A and $P = 4^2 \times 3 = 48\text{ W}$.

Fig. 2.16 (c) shows the situation where both 2$\Omega$ and 3$\Omega$ resistors have been shorted out of the circuit. In this case,

$$R = 1\Omega, I = 12/1 = 12\text{ A and } P = 12^2 \times 1 = 144\text{ W}$$

Because of this excessive current (6 times the normal value), connecting wires and other circuit components can become hot enough to ignite and burn out.

## 'OPENS' IN A SERIES CIRCUIT

In normal series circuit like the one shown in Fig. 2.17 (a), there exists a current flow and the voltage drops across different resistances. If the circuit becomes 'open' anywhere, following two effects are produced :

(i) since 'open' offers infinite resistance, circuit current becomes zero. Consequently, there is no voltage drop across $R_1$ and $R_2$.

(ii) whole of the applied voltage (i.e. 100 V in this case) is felt across the 'open' i.e. across terminals A and B [Fig. 2.17 (b)].

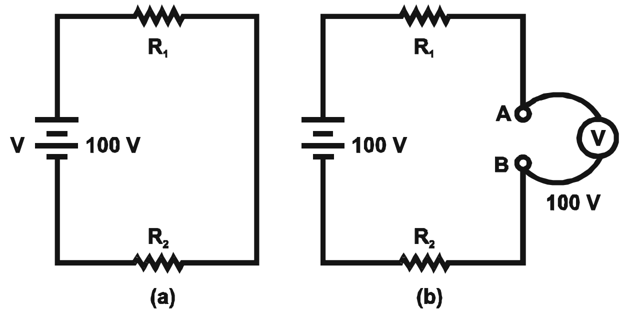



*Fig.2.17*

The reason for this is that $R_1$ and $R_2$ become negligible as compared to the infinite resistance of the 'open' which has practically whole of the applied voltage dropped across it (as per Voltage Divider Rule of art. 1.15). Hence, voltmeter in Fig. 2.17 (b) will read nearly 100 V i.e. the supply voltage.

**'OPENS' IN A PARALLEL CIRCUIT**

Since an 'open' offers infinite resistance, there would be no current in that part of the circuit where it occurs. In a parallel circuit, an 'open' can occur either in the main line or in any parallel branch.

As shown in Fig. 2.18 (a), an open in the main line prevents flow of current to all branches. Hence, neither of the two bulbs glows. However, full applied voltage (i.e. 220 V in this case) is available across the open.

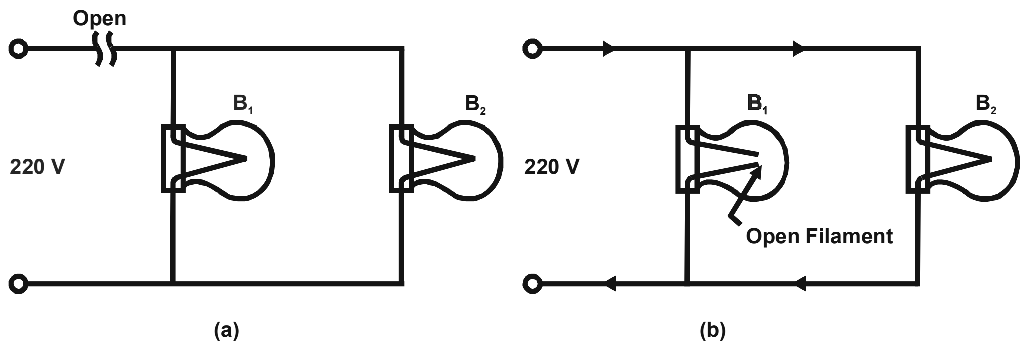



*Fig.2.18*

In this Fig. 2.18 (b), 'open' has occurred in branch circuits of $B_1$. since there is no current in the branch, $B_1$ will not glow. However, as the other bulb remains connected across the voltage supply, it would keep operating normally.

It may be noted that if a voltmeter is connected across the open bulb, it will read full supply voltage of 220 V.

**'SHORTS' IN PARALLEL CIRCUITS**

Suppose a 'short' is placed across $R_3$ (Fig. 2.19). It becomes directly connected across the battery and draws almost infinite current because not only its own resistance but that of the connecting wires AC and BD is negligible. Due to this excessive current, the wires may get hot enough to burn out unless the circuit is protected by a fuse.

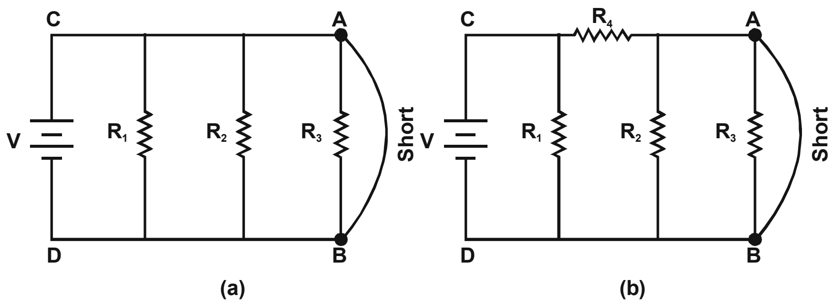



*Fig.2.19*

Following points about the circuit of Fig 2.19 (a) are worth noting :

1. not only is $R_3$ short-circuited but both $R_1$ and $R_2$ are also shorted out i.e. short across one branch means short across all branches.
2. there is no current is shorted resistors. If these were three bulbs, they will not glow.
3. the shorted components are not damaged, For example, if we had three bulbs in Fig. 2.19 (a), they would glow again when circuit is restored to normal conditions by removing the short-circuited.

It may, however, be noted from Fig. 2.19 (b) that a short-circuit across $R_3$ may short out $R_2$ but not $R_1$ since it is protected by $R_4$.

**DIVISION OF CURRENT IN PARALLEL CIRCUITS**

In Fig. 2.20, two resistances are joined in parallel across a voltage V. The current in each branch, as given in Ohm's law, is

$$I_1 = V/R_1 \text{ and } I_2 = V/R_2$$

$$\therefore \quad \frac{I_1}{I_2} = \frac{R_2}{R_1}$$

As $\frac{1}{R_1} = G_1$ and $\frac{1}{R_2} = G_2$

$$\therefore \quad \frac{I_1}{I_2} = \frac{G_1}{G_2}$$

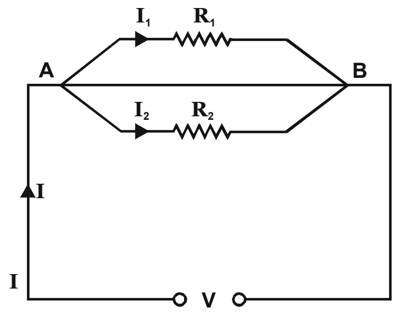



*Fig.2.20*

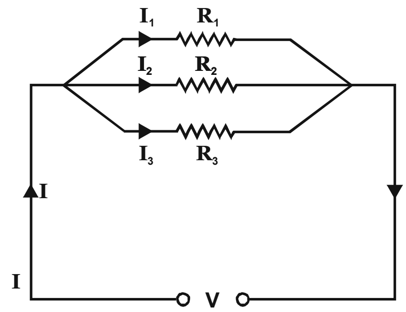



*Fig.2.21*

Hence, the division of current in the branches of a parallel circuit is directly proportional to the conductance of the branches or inversely proportional to their resistances. We may also express the branch currents in terms of the total circuit current thus :

Now $I_1 + I_2 = I; \; \therefore \; I_2 = I - I_1 \; \therefore \; \frac{I_1}{I - I_1} = \frac{R_2}{R_1} \text{ or } I_1R_1 = R_2(I - I_1)$

$$\therefore \quad I_1 = I\frac{R_2}{R_1 + R_2} = I\frac{G_1}{G_1 + G_2} \text{ and } I_2 = I\frac{R_1}{R_1 + R_2} = I\frac{G_2}{G_1 + G_2}$$

This Current Divider Rule has direct application in solving electric circuits by Norton's theorem.

Take the case of three resistors in parallel connected across a voltage V (Fig. 2.21). Total current is $I = I_1 + I_2 + I_3$. Let the equivalent resistance be R. Then

$$V = IR$$

Also $V = I_1R_1 \; \therefore \; IR = I_1R_1$

or $\frac{I}{I_1} = \frac{R_1}{R}$ or $I_1 = IR/R_1$ (i)

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

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

From (i) above, $$I_1 = I\left(\frac{R_2R_3}{R_1R_2 + R_2R_3 + R_3R_1}\right) = I.\frac{G_1}{G_1 + G_2 + G_3}$$

Similarly, $$I_2 = I\left(\frac{R_1R_3}{R_1R_2 + R_2R_3 + R_3R_1}\right) = I.\frac{G_2}{G_1 + G_2 + G_3}$$

$$I_3 = I\left(\frac{R_1R_2}{R_1R_2 + R_2R_3 + R_3R_1}\right) = I.\frac{G_3}{G_1 + G_2 + G_3}$$

**Example 13**. A resistance of 10 $\Omega$ is connected in series with two resistances each of 15 $\Omega$ arranged in parallel. What resistance must be shunted across this parallel combination so that the total current taken shall be 1.5 A with 20 V applied ?

**Sol.** The circuit connections are shown in Fig. 2.22.

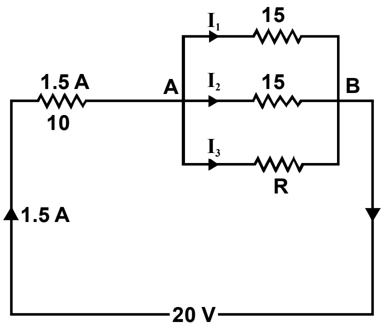



*Fig.2.22*

Drop across 10-$\Omega$ resistor = 1.5 × 10=15 V

Drop across parallel combination, $V_{AB} = 20 - 15 = 5V$

Hence, voltage across each parallel resistance is 5 V.

$$I_1 = 5/15 = 1/3 \text{ A}, I_2 = 5/15 = 1/3 \text{ A}$$

$$I_3 = 1.5 - (1/3 + 1/3) = 5/6 \text{ A}$$

$$\therefore \quad I_3R = 5 \text{ or } (5/6)R = 5 \text{ or } R = 6\Omega$$

**Example 14.** If 20 V be applied across AB shown in Fig. 2.23, calculate the total current, the power dissipated in each resistor and the value of the series resistance to have the total current.

**Sol.** As seen from Fig. 2.23. $R_{AB} = 370/199\,\Omega$

Hence, total current $= 20 \div 370/199 = 10.76\text{ A}$

$$I_1 = 10.76 \times 5(5 + 74.25) = 6.76\text{A}; I_2 = 10.76 - 6.76 = 4\text{A}$$

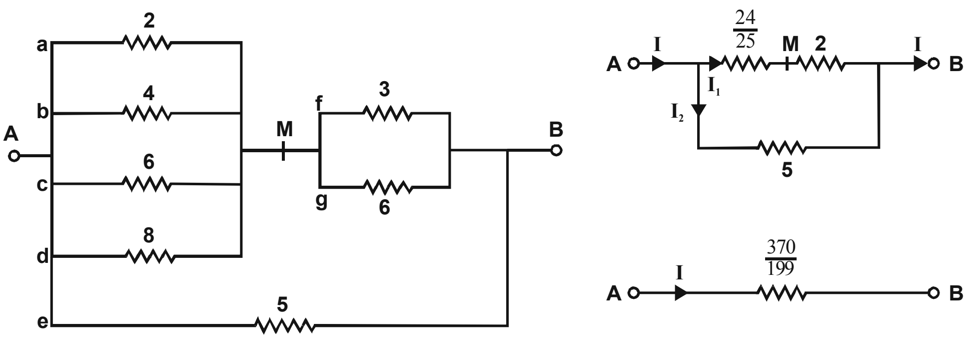



*Fig.2.23*

$$I_f = 6.76 \times 6/9 = 4.51 \text{ A}; I_g = 6.76 - 4.51 = 2.25 \text{ A}$$

Voltage drop across A and M,

$$I_a = V_{AM}/2 = 6.48/2 = 3.24 \text{ A}; I_b = 6.48/4 = 1.62 \text{ A}; I_C = 6.48/6 = 1.08 \text{ A}$$

$$I_d = 6.48/8 = 0.81 \text{ A}, I_e = 20/5 = 4 \text{ A}$$

**POWER DISSIPATION**

$$P_a = I_a^2 R_a = 3.24^2 \times 2 = 21\text{W}, {}_b = 1.62^2 \times 4 = 10.4 \text{ W}, P_c = 1.08^2 \times 6 = 7 \text{ W}$$

$$P_d = 0.81^2 \times 8 = 5.25 \text{ W}, P_e = 4^2 \times 5 = 80\text{W}, P_f = 4.51^2 \times 3 = 61 \text{ W}$$

$$P_g = 2.25^2 \times 6 = 30.4 \text{ W}$$

The series resistance required is 370/199 $\Omega$

Incidentally, total power dissipated = $I^2 R_{AB} = 10.76^2 \times 370/199 = 215.3 \text{ W}$ (as a check).

**Example 15.** Calculate the values of different currents for the circuit shown in Fig, 2.24. What is the total circuit conductance ? and resistance ?

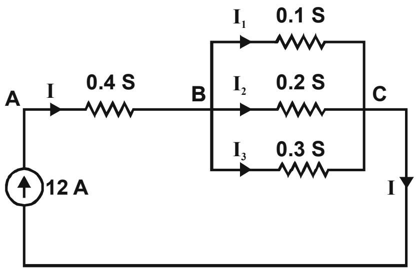



*Fig.2.24*

**Sol.** As seen, $I = I_1 + I_2 + I_3$. The current division takes place at point B.

As seen

$$I_1 = I.\frac{G_1}{G_1 + G_2 + G_3}$$

$$= 12 \times \frac{0.1}{0.6} = 2\text{A}$$

$$I_2 = 12 \times 0.2/0.6 = 4\text{A}$$

$$I_3 = 12 \times 0.3/0.6 = 6\text{A}$$

$$G_{BC} = 0.1 + 0.2 + 0.3 = 0.6 \text{ S}$$

$$\frac{1}{G_{AC}} = \frac{1}{G_{AB}} + \frac{1}{G_{BC}} = \frac{1}{0.4} + \frac{1}{0.6} = \frac{25}{6}\text{S}^{-1}$$

$$\therefore \quad R_{AC} = 1/G_{AC} = 25/6\,\Omega$$

$$G_{AC} = \frac{6}{25}S$$

**Example 16.** Compute the values of three branch currents for the circuits of Fig. 2.25 (a). What is the potential difference between points A and B ?

**Sol.** The two given current sources may be combined together as shown in Fig. 2.25 (b).

Net Current = 25-6=19 A because the two currents flow in opposite directions.

Now, $G = 0.5 + 0.25 + 0.2 = 0.95; \quad I_1 = I\frac{G_1}{G} = 19 \times \frac{0.5}{0.95} = 10 \text{ A}$

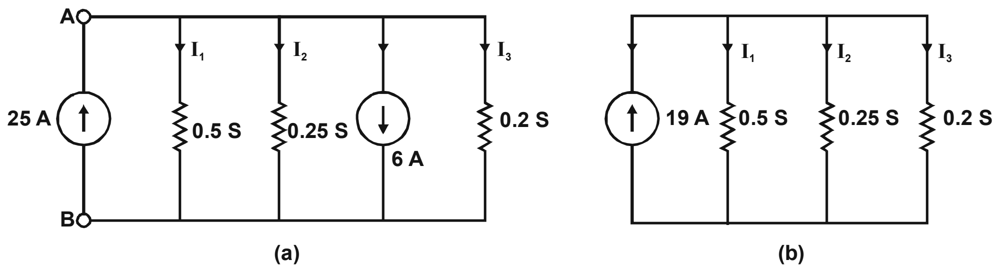



*Fig.2.25*

$$I_2 = I\frac{G_2}{G} = 19\times\frac{0.5}{0.95} = 5\text{ A};\ I_3 = I\frac{G_3}{G} = 19\times\frac{0.2}{0.95} = 4\text{A}$$

$$V_{AB} = I_1R_1 = \frac{I_1}{G_1} = \frac{I_2}{G_2} = \frac{I_3}{G_3} \quad \therefore\ V_{AB} = \frac{10}{0.5} = 20\text{ V}$$

The same voltage acts across the three conductances.

**Example 17.** Two conductors, one of copper and the other of iron, are connected in parallel and at 20°C carry equal currents. What proportion of current will pass through each if the temperature is raised to 100°C ? Assume $\alpha$ for copper as 0.0042 and for iron as 0.006 per °C at 20°C. Find also the values of temperature coefficients at 100°C.

**Sol.** Since they carry equal current at 20°C, the two conductors have the same resistances at 20°C i.e. $R_{20}$. As temperature is raised, their resistances increase unequally.

For Cu, $R_{100} = R_{20}(1+80\times0.0042) = 1.336\,R_{20}$

For iron $R'_{100} = R_{20}(1+80\times0.006) = 1.48\,R_{20}$

As seen from Art. 1.25, current through Cu conductor is

$$I_1 = I\times\frac{R'_{100}}{R_{100}+R'_{100}} = I\times\frac{1.48\,R_{20}}{2.816\,R_{20}} = 0.5256\,I \text{ or } 52.56\% \text{ of I}$$

Hence, current through Cu conductor is 52.56 per cent of the total current. Obviously, the remaining current i.e. 47.44 per cent passes through iron.

Or current through iron conductor is

$$I_2 = I\times\frac{R'_{100}}{R_{100}+R'_{100}} = I\times\frac{1.336\,R_{20}}{2.816\,R_{20}} = 0.4744\,I \text{ or } 47.44\% \text{ of I}$$

For Cu, $\alpha_{100} = \dfrac{1}{(1/0.0042)+80} = 0.00314°C^{-1}$

For iron, $\alpha_{100} = \dfrac{1}{(1/0.006)+80} = 0.0040°C^{-1}$

**Example 18.** A battery of unknown e.m.f. is connected across resistances as shown in Fig. 2.26. The voltage drop across the 8-$\Omega$ resistor is 20 V. What will be the current reading in the ammeter ? What is the e.m.f. of the battery?

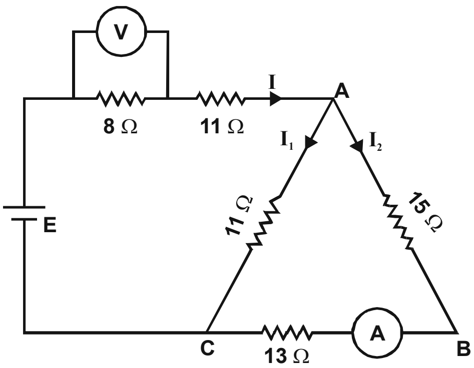



*Fig.2.26*

**Sol.** Current through 8-$\Omega$ resistance = 20/8=2.5 **A**

This current is divided into two parts at point A; one part going along path AC and the other along path ABC which has a resistance of 28 $\Omega$.

$$I_2 = 2.5 \times \frac{11}{(11+28)} = 0.7$$

Hence, ammeter reads 0.7 A.
Resistance between A and C = (28×11/39) ohm.
Total circuit resistance = 8+11+(308/39) = 1049/39) $\Omega$

$$\therefore \quad E = 2.5 \times 1049/39 = 67.3\text{ V}$$

**EQUIVALENT RESISTANCE**

The equivalent resistance of a circuit (or network) between its any two points (or terminals) is given by that single resistance which can replace the entire given circuit between these two points. It should be noted that resistance is always between two given points of a circuit and can have different values for different point-pairs as illustrated by Example 2.6. it can usually be found by using series and parallel laws of resistances. Concept of equivalent resistance is essential for understanding network theorems like Thevenin's theorem and Norton's theorem etc.

**Example 19.** Find the equivalent resistance of the circuit given in Fig. 2.27 (a) between the following points (i) A and B (ii) C and D (iii) E and F (iv) A and F and (v) A and C. Numbers represent resistances in ohm.

**Solution. (i) Resistance Between A and B**

In this case, the entire circuit to the right side of AB is in parallel with 1$\Omega$ resistance connected directly across points A and B.

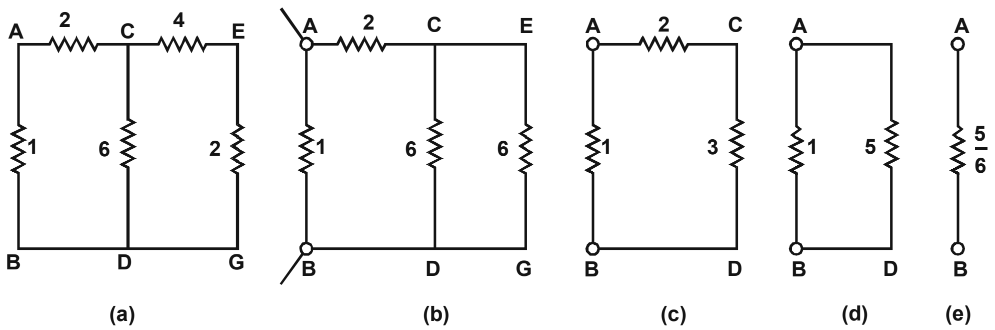



*Fig.2.27*

As seen, there are two parallel paths across points C and D; one having a resistance of 6 $\Omega$ and the other of (4 + 2) = 6 $\Omega$. As shown in Fig. 2.27 (c), the combined resistance between C and D is = 6 || 6 = 3 $\Omega$. Further simplifications are shown in Fig. 2.27 (d) and (e). As seen, $R_{AB} = 5/6\ \Omega$.

**(ii) Resistance between C and D**

As seen from Fig. 2.27 (a), there are three parallel paths between C and D (i) CD itself of 6$\Omega$ (ii) CEFD of (4 + 2)= 6 $\Omega$ and (iii) CABD of (2 + 1) = 3 $\Omega$. It has been shown separately in Fig. 2.28 (a). The equivalent resistance $R_{CD}$= 3 || 6 || 6 = 1.5 $\Omega$ as shown in Fig. 2.28 (b).

**(iii) Resistance between E and F**

In this case, the circuit to the left side of EF is in parallel with the 2 $\Omega$ resistance connected directly across E and F. This circuit consists of a 4 $\Omega$ resistance connected in series with a parallel circuit of 6|| (2 + 1) = 2 $\Omega$ resistance. After various simplifications as shown in Fig. 2.29, $R_{EF}$=2 || 6=1.5 $\Omega$.

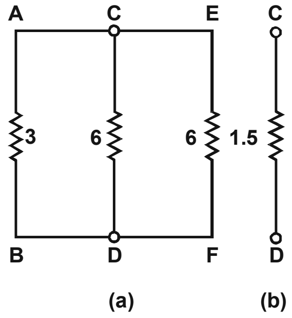



*Fig.2.28*

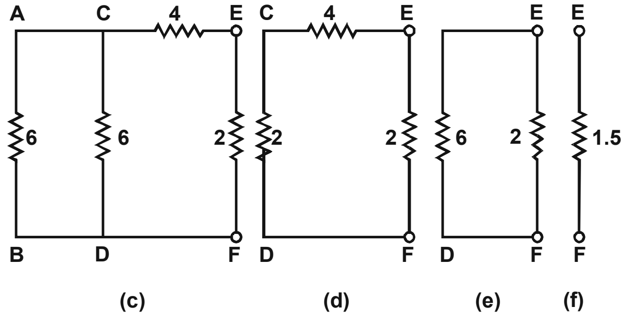



*Fig.2.29*

**(iv) Resistance Between A and F**

As we go from A and F, there are two possible routes to begin with : one along ABDF and the other along AC. At point C, there are again two alternatives, one along CDF and the other along CEF.

As seen from Fig.2.30 (b), $R_{CD} = 6 \parallel 6 = 3\ \Omega$ . Further simplification of the original circuit as shown in Fig. 2.30 (c), (d) and (e) gives $R_{AF} = 5/6\ \Omega$ .

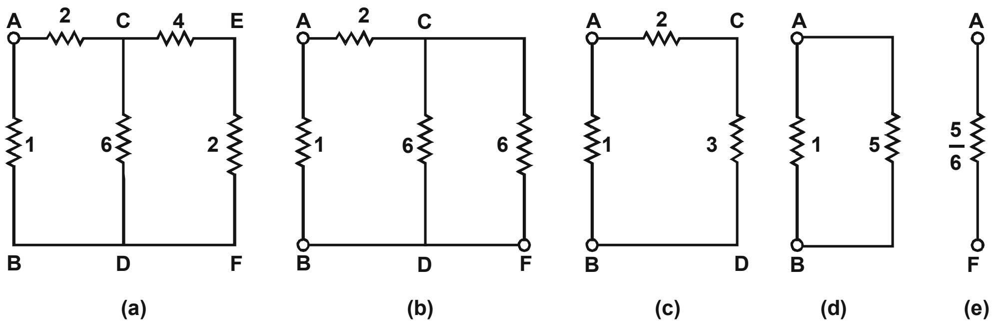



*Fig.2.30*

**(v) Resistance Between A and C**

In this case, there are two parallel paths between A and C; one is directly from A to C and the other is along ABD. At D, there are again two parallel paths to C; one is directly along DC and the other is along DFEC.

As seen from Fig. 2.31 (b), $R_{CD} = 6 \parallel 6 = 3\ \Omega$ . Again, from Fig. 2.31 (d), $R_{AC} = 2 \parallel 4 = 4/3\ \Omega$ .

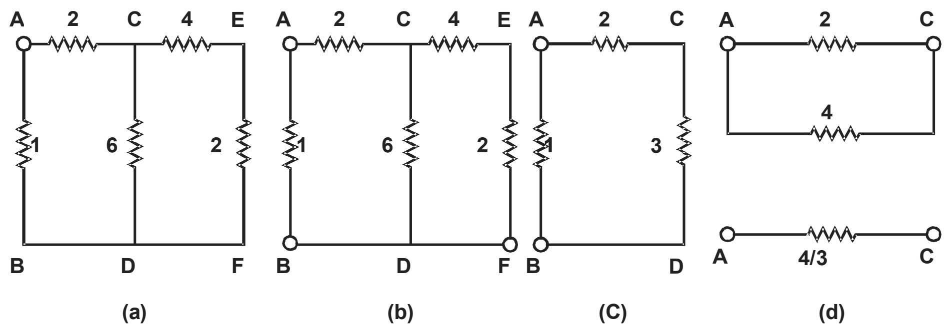



*Fig.2.31*

**Example 20.** Two resistors of values 1 k$\Omega$ and 4 $\Omega$ are connected in series across a constant voltage supply of 100V. A voltmeter having an internal resistance of 12 k$\Omega$ is connected across the 4 k$\Omega$ resistor. Draw the circuit and calculate.

(a) true voltage across 4 k$\Omega$ resistor before the voltmeter was connected.

(b) actual voltage across 4 k$\Omega$ resistor after the voltmeter is connected and the voltage recorded by the voltmeter.

(c) change in supply current when voltmeter is connected.

(d) percentage error in voltage across 4 k$\Omega$ resistor.

**Sol.** (a) True voltage drop across 4 k$\Omega$ resistor as found by voltage-divider rule is $100 \times 4/5 = 80$ V

Current from the supply = $100/(4 + 1) = 20$ mA

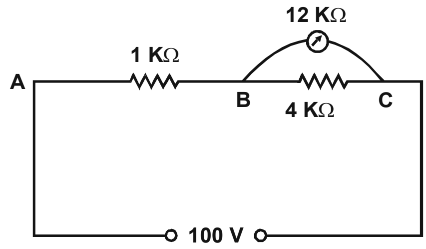



*Fig.2.32*

(b) In Fig. 2.32, voltmeter has been joined across the 4 k$\Omega$ resistor. The equivalent resistance between B and C

$$= 4 \times 12/16 = 3\ \text{k}\Omega$$

Drop across B and C $= 100 \times 3/(3+1) = 75\ \text{V}$.

(c) Resistance between A and C = 3 + 1 = 4 k$\Omega$

New supply current = 100/4=25 mA

$\therefore$ increase in current = 25 - 20 = 5 mA

(d) Percentage error in voltage $= \left(\dfrac{\text{actual voltage - true voltage}}{\textit{true}\ \text{voltage}}\right) \times 100 = \dfrac{(75-80)}{80} \times 100 = -6.25\%$

The reduction in the value of voltage being measured is called voltmeter loading effect because voltmeter loads the circuit element across which it is connected. Smaller the voltmeter resistance as compared to the resistance across which it is connected, greater the loading effect and hence, greater the error in the voltage reading. Loading effect cannot be avoided but can be minimized by selecting a voltmeter of resistance much greater than that of the network across which it is connected.

**Example 21.** In the circuit of Fig. 2.33, find the value of supply voltage V so that 20-$\Omega$ resistor can dissipate 180 W.

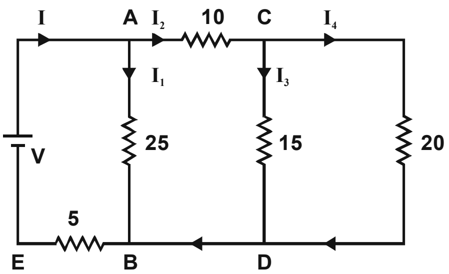



*Fig.2.33*

**Sol.** $I_4^2 \times 20 = 180\ \text{W};\quad I_4 = 3\ \text{A}$

Since 15 $\Omega$ and 20 $\Omega$ are in parallel,

$$I_3 \times 15 = 3 \times 20 \qquad \therefore \quad I_3 = 4\ \text{A}$$

$$I_2 = I_3 + I_4 = 4 + 3 = 7\text{A}$$

Now, resistance of the circuit to the right of point A is

$$= 10 + 15 \times 20/35 = 130/7\ \Omega$$

$\therefore$ $I_1 \times 25 = 7 \times 130/7$

$\therefore$ $I_1 = 26/5\ \text{A} = 5.2\ \text{A}$

$\therefore$ $I = I_1 + I_2 = 5.2 + 7 = 12.2\ \text{A}$

Total circuit resistance

$R_{AE} = 5 + 25 \parallel 130/7 = 955/61\ \Omega$

$\therefore \qquad V = I.R_{AE} = 12.2 \times 955/61 = 191\,V$

**Example 22.** For the simple ladder network shown in Fig. 2.34, find the input voltage $V_i$ which produces a current of 0.25 A in the 3 - $\Omega$ resistor. All resistances are in ohm.

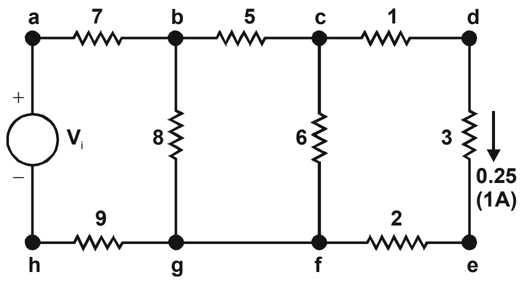



*Fig.2.34*

**Sol.** We will assume a current of 1 A in the 3-$\Omega$ resistor. The voltage necessary to produce 1A bears the same ratio to 1 A as $V_i$ does to 0.25 A because of the linearity of the network. I is known as Current Assumption technique.

Since $\quad R_{cdef} = R_{cf} = 6\Omega$

Hence, $\quad I_{cf} = 1\,A$

and $\quad V_{cf} = V_{cdef} = 1 \times 6 = 6V.$

Also, $\quad I_{bc} = 1 + 1 = 2\,A$

$$V_{bg} = V_{bb} + V_{ef} = 2 \times 5 + 6 = 16V$$

$$I_{bg} = 16/8 = 2A$$

$$I_{ab} = I_{bc} + b_{bg} = 2 + 2 = 4A$$

$$V_i = V_{ab} = V_{bg} + V_{gh}$$

$$= 4 \times 7 + 16 + 4 \times 9 = 80V$$

Taking the proportion, we get

$$\frac{80}{1} = \frac{V_i}{0.25} \quad \therefore \; V_i = 80 \times 0.25 = 20V$$

**Example 23.** In this circuit of Fig. 2.35, find the value $R_1$ and $R_2$ so that $I_2 = I_1/n$ and the input resistance as seen from points A and B is R ohm.

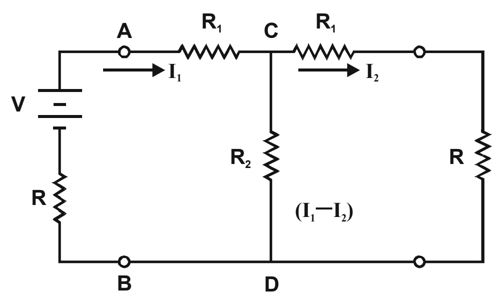



*Fig.2.35*

**Sol.** As seen, the current through $R_2$ in $(I_1 - I_2)$. Hence, p.d. across points C and D is

$$R_2(I_1 - I_2) = (R_1 + R)I_2 \text{ or } R_2 I_1 = (R_1 + R_2 + R)I_2$$

$$\therefore \quad \frac{I_1}{I_2} = \frac{R_1 + R_2 + R}{R_2} = n \qquad \text{(i)}$$

The input resistance of the circuit as viewed from terminals A and B is required to be R.

$$\therefore \quad R = R_1 + R_2 \parallel (R_1 + R)$$

$$= R_1 + \frac{R_1 + R}{n} \qquad \text{using Eq. (i)}$$

$$R(n-1) = R_1(n+1)$$

$$\therefore \quad R_1 = \frac{n-1}{n+1} R \text{ and } R_2 = \frac{R_1 + R}{(n-1)} = \frac{2n}{n^2 - 1} R$$

## RESISTOR COLOUR CODE

The resistance value of any resistor can be measured by using an ohmmeter. But this is seldom necessary. Most wire-wound resistors have their resistance value in ohms printed on the body of the resistor. Many carbon resistors are similarly marked, but are often mounted in such a manner that it is difficult or impossible to read the resistance value. Additionally, heat often discolors the resistor body, making the printed marking illegible, and many carbon resistors are so small that a printed marking cannot be use. Thus, a color code marking is used to identify the resistance value of carbon resistors.

There is only one color code for carbon resistors, but there are two systems or methods used to paint this color code on resistors. One is the body-end-dot system, and the other is the end-to-center band system.

In each color code system, three colors are used to indicate the resistance value in ohms, and a fourth color is sometimes used to indicate the tolerance of the resistor. By reading the colors in the correct order and by substituting numbers from the color code, the resistance value of a resistor can be determined.

It is very difficult to manufacture a resistor to an exact standard of ohmic values. Fortunately, most circuit requirements are not extremely critical. For many used the actual resistance in ohms can be 20 percent higher or lower than the value marked on the resistor without causing difficulty. The percentage variation between the marked value and the actual value of a resistor is known as the "tolerance" of a resistor. A resistor coded for a 5 per cent tolerance will not be more than 5 percent higher or lower than the value indicated by the color code.

| RESISTOR COLOR CODE | | |
|---|---|---|
| COLOR | NUMBER | TOLERANCE |
| BLACK | 0 | - - - - |
| BROWN | 1 | 1% |
| RED | 2 | 2% |
| ORANGE | 3 | 3% |
| YELLOW | 4 | 4% |
| GREEN | 5 | 5% |
| BLUE | 6 | 6% |
| VIOLET | 7 | 7% |
| GRAY | 8 | 8% |
| WHITE | 9 | 9% |
| GOLD | - - - - | 5% |
| SILVER | - - - - | 10% |
| NO COLOR | - - - - | 20% |

*Fig.2.36, Resistance color code.*

The resistor color code (see figure 2.36) is made up of a group of colors, numbers, and tolerance values. Each color is represented by a number and in most cases by a tolerance value.

When the color code is used with the end-to-center band marking system, the resistor is normally marked with bands of color at one end of the resistor. The body or base color of the resistor. The body or base color of the resistor has nothing to do with the color code, and in no way indicates a resistance value. To prevent confusion, this body will never be the same color as any of the bands indicating resistance value.

When the end-to-center band marking system is used, the resistor will be marked by either three or four bands. The first color band (nearest to the end of the resistor) will indicate the first digit in the numerical resistance value. This band will never be gold or silver in color.

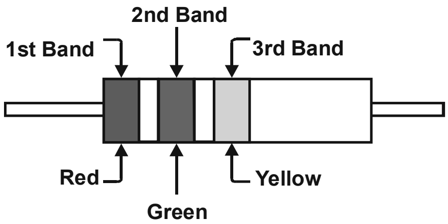



*Fig.2.37, End-to-center band marking.*

The second color band (refer to figure 2.37) will always indicate the second digit of ohmic value. It will never be gold or silver in color. The third color band indicates the number of zeroes to be added to the two digits derived from the first and second bands, except in the following two cases :

1. If the third band is gold in color, the first two digits must be multiplied by 10 percent.

2. If the third band is silver in color, the first two digits must be multiplied by 1 percent.

If there is a fourth color band, it is used as a multiplier for percentage of tolerance, as indicated in the color code chart in figure 2.37. If there is no fourth band, the tolerance is understood to be 20 percent.

Figure 2.37 illustrates the rules for reading the resistance value of a resistor marked with the end-to-center band system. This resistor is marked with three bands of color, which must be read from the end towards the center.

These are values that should be obtained :

**Numerical**

| Color | Value | Significance |
|---|---|---|
| 1st band - Red | 2 | 1st digit |
| 2nd band-Green | 5 | 2nd digit |
| 3rd band -Yellow | 4 | No. of zeroes to add |

There is no fourth color band, so the tolerance is understood to be 20 percent. 20 percent of 250,000=50,000.

Since the 20 percent tolerance is plus or minus.

Maximum resistance = 250,000+50,000
= 300,000 ohms
Minimum resistance = 250,000 - 50,000
= 200,000 ohms.

Figure 2.38 contains a resistor with another set of colors. This resistor code should be read as follows :


*Fig. 2.38, Resistor color code example.*

The resistance of this resistor is 86,000 $\pm$ 10 percent ohms. The maximum resistance is 94,600 ohms and the minimum resistance is 77,400 ohms.

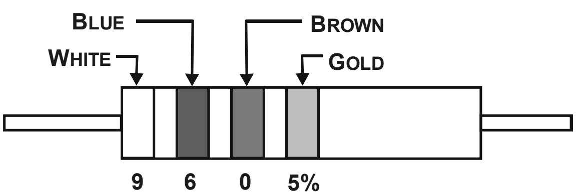



*Fig. 2.39, Resistor color code example.*

As another example, the resistance of the resistor in figure 2.39 is 960 $\pm$ 5 percent ohms. The maximum resistance is 1,008 ohms, and the minimum resistance is 912 ohms.

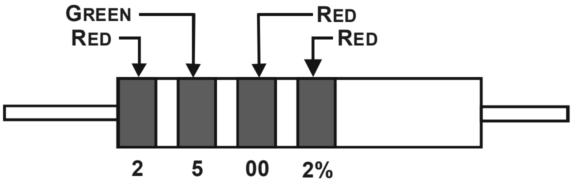



*Fig. 2.40, Resistor with 2 percent tolerance.*

Sometimes circuit considerations dictate that the tolerance must be smaller than 20 percent. Figure 2.40 shows an example of a resistor with a 2 percent tolerance. The resistance value of this resistor is 2,500 $\pm$ 2 percent ohms. The maximum resistance is 2,550 ohms, and the minimum resistance is 2,450 ohms.

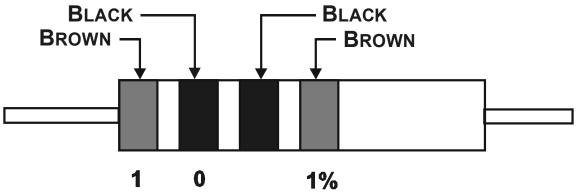



*Fig. 2.41, Resistor with black third color band.*

Figure 2.41 contains an example of a resistor with a black third color band. The color code value of black is zero, and the third band indicates the number of zeroes to be added to the first two digits.

In this case, a zero number of zeroes must be added to the first two digits; therefore, no zeroes are added. Thus, the resistance value is 10 $\pm$ 1 percent ohms. The maximum resistance is 10.1 ohms, and the minimum resistance is 9.9 ohms.

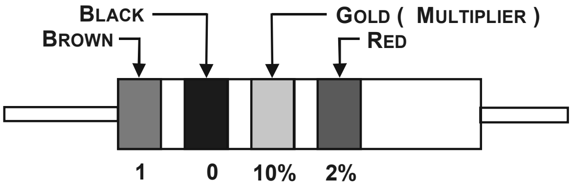



*Fig. 2.42, Resistor with a gold third band.*

There are two exceptions to the rule stating the third color band indicates the number of zeroes. The first of these exceptions is illustrated in figure 2.42. When the third band is gold in color, it indicates that the first two digits must be multiplied by 10 percent. The value of this resistor is

$$10 \times .10 \pm 2\% = 1 \pm .02 \text{ ohms}$$

When the third band is silver, as is the case in figure 2.43, the first two digits must be multiplied by 1 percent. The value of the resistor is .45 $\pm$ 10 percent ohms.

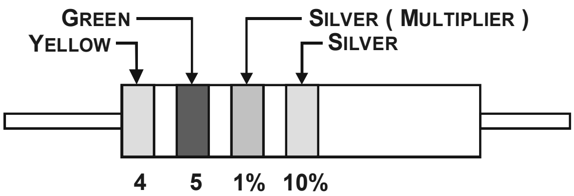



*Fig. 2.43, Resistor with a silver third band.*

**BODY-END-DOT SYSTEM**

The body-end-dot system of marking is rarely used today. A few examples will explain it. The location of the colors has the following significance :

Body color .................1st digit of ohmic value
End color ....................2nd digit of ohmic value
Dot color ....................Number of zeroes to be added

If only one end of the resistor is painted, it indicates the second figure of the resistor value, and the tolerance will be 20 percent. The other two tolerance values are gold (5 percent) and silver (10 percent). The opposite end of the resistor will be painted to indicate a tolerance other than 20 percent. Figure 2.44 shows a resistor coded by the body-end-dot system.

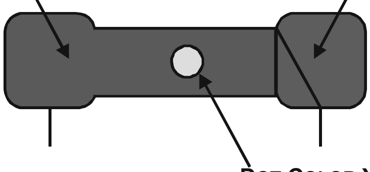



***Fig. 2.44, Resistor coded with body-end-dot system.***

The values are as follows :

Body - 1st digit - 2
End - 2nd digit - 5
Dot - No. of zeroes - 0000 (4).

The resistor value is 250,000 $\pm$ 20 percent ohms. The tolerance is understood to be 20 percent because no second dot is used.

If the same color is used more than once, the body, end, and dot may all be the same color, or any two may be the same ; but the color code is used in exactly the same way. For example, a 33,000 ohm resistor will be entirely orange.

■■■

# CHAPTER : 3
# KIRCHHOFF'S LAW

**GENERAL**

Many laws and theorems are available for solving networks of conductors both active and passive. The main advantage of these theorems is that they save time and considerably reduce the laborious mathematical work involved in solving networks thereby minimising chances of error. The different laws and theorems discussed in the text are :

1. Kirchhoff's Laws
2. Maxwell's Loop Current Theorem
3. Superposition Theorem
4. Thevenin's Theorem
5. Norton's Theorem
6. Maximum Power Transfer Theorem
7. Delta/Star Transformation.
8. Star/Delta Transformation.

**KIRCHHOFF'S LAWS**

These laws are more comprehensive than Ohm's law and are used for solving electrical networks which may not be readily solved by the latter. The two laws are :

**Kirchhoff's First Law or Point Law or Current Law (KCL)**

It states that :

In any network of conductors, the algebraic sum of the currents meeting at a point (or junction) is zero.

Put in another way, it simply means that the total current leaving a junction is equal to the total current entering the junction. It is obviously true because there is no accumulation or depletion of current at any junction of the network

**Explanation**

Consider the case of a few conductors meeting at a point A as in Fig. 3.1. Since two conductors have currents leading to point A whereas others have currents leading away from point A. Assuming the incoming currents to be positive and the outgoing currents negative, we have

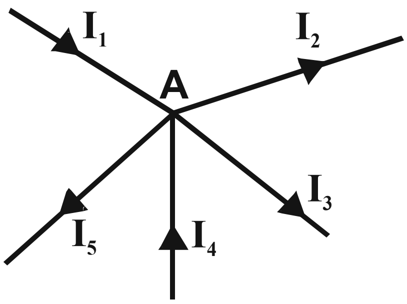



*Fig. 3.1*

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

or $\quad I_1 + I_4 - I_2 - I_3 - I_5 = 0$

or $\quad I_1 + I_4 = I_2 + I_3 + I_5$

or $\quad$ incoming currents = outgoing currents

We can express the above conclusion thus

$\Sigma I = 0$ .......at a junction

**Kirchhoff's Second Law or Mesh Law or Voltage Law (KVL)**

It states that :

The algebraic sum of the product of current and resistance in each of the conductors in any closed mesh (or path) in a network plus the algebraic sum of the e.m.f.s in that path is zero.

$\Sigma IR + \Sigma$ e.m.$f.$ = 0 ...........round a mesh

The basis of this law is this : If one starts from a particular junction and goes round the mesh till one comes back to the starting point, then one must be at the same potential with which one started. Hence, it means that all the sources of e.m.f met on the way must necessarily be equal to the voltage drops in the resistances, every voltage being given its proper sign, plus or minus.

**Determination of Sign**

In applying Kirchhoff's laws to specific problems, particular attention should be paid to the algebraic signs of voltage drops and e.m.f.s, otherwise results will come out to be wrong. Following sign convention is suggested:

A rise in voltage should be given a +ve sign and a fall in voltage a -ve sign. Keeping this in mind, it is clear that as we go from the -ve terminal of a battery to its +ve terminal, there is a rise in potential, hence this voltage should be given a + ve sign [Fig. 3.2 (a)]. If on the other hand, we go from +ve terminal to -ve terminal [Fig.3.2 (b)], then there is a fall in potential, hence this voltage should be preceded by a -ve sign. It is important to note that the sign of the battery e.m.f. is independent of the direction of the current through that branch.

Now, take the case of a resistor. If we go through a resistor in the same direction as the current, then there is a fall in potential because current flows from higher to lower potential. Hence, this voltage fall should be taken -ve. However, if we go in a direction opposite to that of the current, then there is a rise in voltage. Hence, this voltage rise should be given a positive sign.

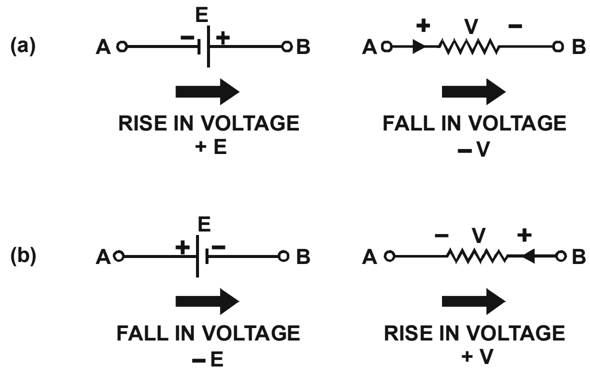



*Fig.3.2*

It is clear that the sign of voltage drop across a resistor depends on the direction of current through that resistor.

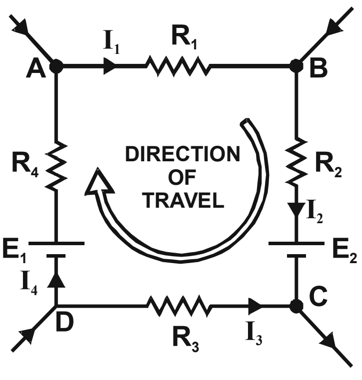



*Fig.3.3*

Consider the closed path ABCDA in Fig. 3.3. Different voltage drops will have the following signs :

$I_1\ R_1$ *is* -ve (fall in potential)
$I_2\ R_2$ *is* -ve (fall in potential)
$I_3\ R_3$ *is* +ve (rise in potential)
$I_4\ R_4$ *is* -ve (fall in potential)
$E_2$ is -ve (fall in potential)
$E_1$ is +ve (rise in potential)

Using Kirchhoff's Second Law,
we get

$$-I_1 R_1 - I_2 R_2 + I_3 R_3 - I_4 R_4 - E_2 + E_1 = 0$$

or $$I_1 R_1 + I_2 R_2 - I_3 R_3 + I_4 R_4 = E_1 - E_2$$

**Assumed Direction of Current**

In applying Kirchhoff's laws to electrical networks, the question of assuming proper direction of current usually arises. The direction of current flow may be assumed either clockwise or anticlockwise. If the assumed direction of current is not the actual direction, then on solving the question, this current will be found to have a minus sign. If the answer is positive, then assumed direction is the same as actual direction. However, the important point is that once a particular direction has been assumed, the same should be used throughout the solution of the question.

**Note.** It should be noted that Kirchhoff's laws are applicable to both d.c. and a.c. voltages and currents. However, in the case of alternating currents and voltages, any e.m.f. of self-inductance or that existing across a capacitor should also be taken into account.

1. A bridge network ABCD is arranged as follows: resistances between terminals A-B, B-C, C-D, D-A and B-D are 10, 30, 15, 20, and 40 Ω respectively. A 2 V battery of negligible internal resistance is connected between terminals A and C. Determine the value and direction of the current in 40 Ω resistor.

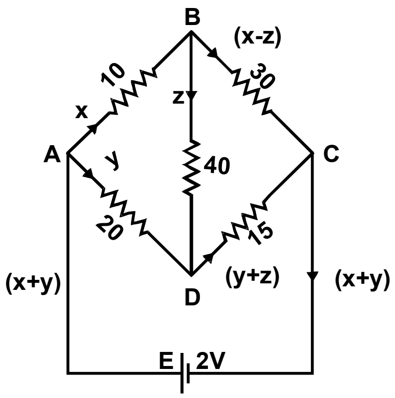



*Fig.3.4*

***Solution.*** The directions of various currents are as shown in Fig. 3.4. Applying Kirchhoff's Second Law to various closed circuits, we have

**Circuit ABDA**

$-10x - 40z + 20y = 0$

or $x - 2y + 4z = 0$ . (i)

**Circuit BCDB**

$-30(x - z) + 15(y + z) + 40z = 0$

$6x - 3y - 17z = 0$ (ii)

**Circuit ADCEA**

$-20y - 15(y + z) + 2 = 0$

$35y + 15z = 2$ (iii)

Multiplying Eq. (i) by 6 and subtracting Eq. (ii) from it, we have $9y - 41z = 0$ (iv)

Multiplying Eq. (iii) by 9 and then subtracting Eq. (iv) from it after multiplying it by 35, we have,

$1570z = 18, \quad z = 18/1570 = 9/785$ A.

Since z turns out to be positive, its actual direction of flow is the same as assumed in Fig. 3.4.

2 Find the value of R and the current through it in the circuit shown in Fig. 3.5 (a) below when the current is zero in branch OA.

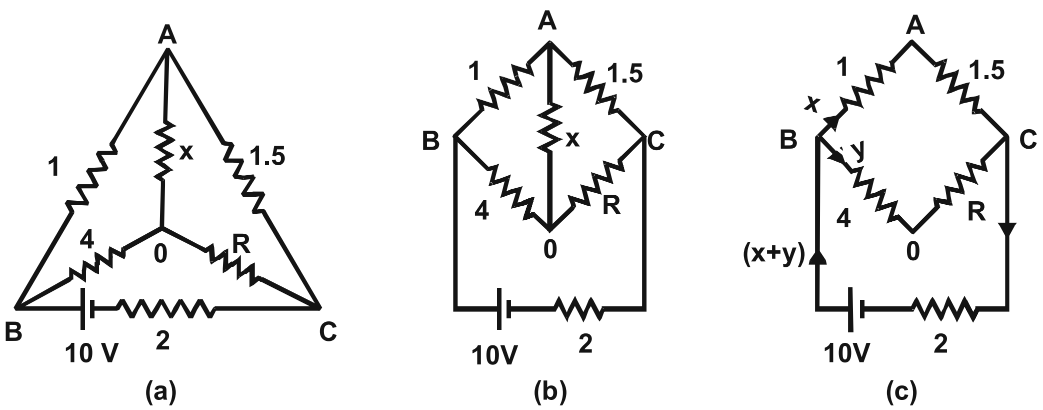



*Fig.3.5*

***Solution.*** The given circuit can be redrawn as shown in Fig. 3.5 (b). As seen, it is nothing else but a wheatstone bridge circuit.

Since there is no current in branch OA, it means that the given bridge is balanced.

Under balanced conditions, the products of the resistances of the opposite arms of such a bridge are equal.
$R \times 1 = 4 \times 1.5$ or $R = 6\,\Omega$
It is given that there is no current in the branch *OA* . In that case, the resistance *X* of the branch *OA* can be removed as shown in Fig. 3.5 (c) without, in any way, electrically affecting the circuit.

Now, there are two parallel paths between points *B* and *C* of resistance (6 + 4) = 10 Ω and (1 + 1.5) = 2. 5 Ω

Total resistance between points *B* and *C* is
$= 10 \times 2.5 / (10 + 2.5) = 2\,\Omega$
Total circuit resistance $= 2 + 2 = 4\,\Omega$
Circuit current $= 10/4 = 2.5$ A

This current divides into two parts at point *B,* one part going along *BAC* and the other along *BOC.*
Current going along path *BOC* is y= 2.5/12.5 = 0.5 A
The same current, obviously, passes through the given resistance R.

3. Two batteries A and B are connected in parallel and a load of 10 Ω is connected across their terminals. A has an e.m.f of 12 V and internal resistance of 2Ω ; B has an e.m.f. of 8 V and an internal resistance of 1Ω. Using Kirchhoff's laws, determine the values and direction of the currents flowing in each of the batteries and in the external resistance. Also, determine the potential difference across the external resistance.
***Solution.*** The two batteries are connected in parallel as shown in Fig.3.6. It should be noted that unless stated otherwise, the similar ends of the two batteries are assumed to be joined together . Let the directions of the two branch currents be as shown.

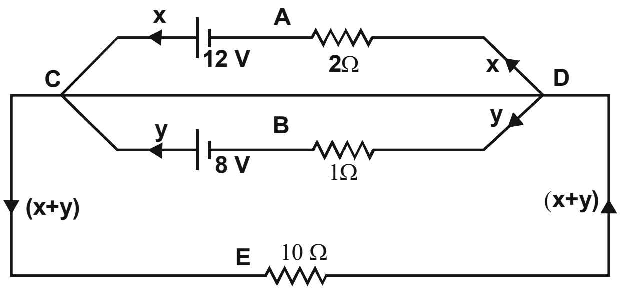



*Fig.3.6*

Applying Kirchhoff's laws to the two loops, we get the following equations:-
**Loop CEDAC**. Starting from point C anticlockwise, we get
$-10\,(x+y) - 2x + 12 = 0$ or $6x + 5y = 6$ (i)

**Loop CEDBC**. Starting from point C, we get
$-10(x + y) - 1 \times y + 8 = 0$ or $10x + 11y = 8$ (ii)

Solving for x and y from (i) and (ii) we get
$x = 13/8 = 1.625$A (discharge), $y = 0.75$ A (charge)
Since y turns out to be *negative,* its actual direction of flow is opposite to that shown in Fig. 3.6 . It is a *charging* current and not discharging one.
Current in 10 Ω resistor = (1.625 - 0.75) = 0.875 A
P. D. across external resistor = 10 x 0.875 =8.75 V

4. *Determine the current x in the 4-ohm resistance in the circuit shown in Fig. 3.7*(a) below.
***Solution.*** The assumed distribution of currents is shown in Fig. 3.7 (b). Applying Kirchhoff's laws to different closed loops, we get
**Circuit EFADE**
$-2y + 10z + 1\,(x-y-6) = 0$
or $x - 3y + 10z = 6$ (i)

**Circuit ABCDA**
$-2(y+z+6) - 10 + 3\,(x-y-z-6) - 10z = 0$
or $3x - 5y - 15z = 40$ (ii)

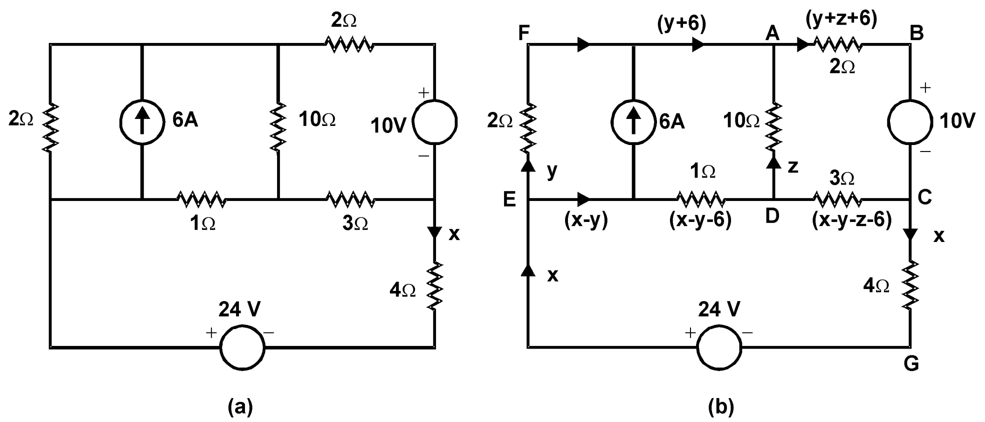



*Fig.3.7*

**Circuit EDCGE**

$-1(x-y-6)-3(x-y-z-6)-4x+24=0$

or $8x-4y-3z=48$ (iii)

Multiplying Eq. (i) by 5 and Eq. (ii) by 3 and then subtracting Eq. (ii) from Eq. (i), we get

$-4x+95z=-90$

or $4x-95z=90$ (iv)

Next, multiplying Eq. (ii) by 4 and Eq. (iii) by 5 and subtracting Eq. (iii) from Eq. (ii), we get

$-28x-45z=-80$

or $28x+45z=80$ (v)

Multiplying Eq. (iv) by 45 and Eq. (v) by 95 and adding the two, we get

$284x=1165$ or $x=1165/284=4.1$ A

5. *Formulate the Kirchhoff voltage law equations for the circuit of Fig.3.8 and find the values of $I_1$, $I_2$ and $I_s$.*

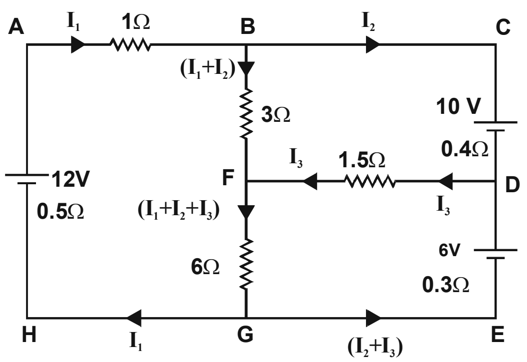



*Fig.3.8*

***Solution.*** **Circuit ABFGHA**

$-I_1 \times 1 - 3(I_1+I_2) - 6(I_1+I_2+I_3) - 0.5I_1 + 12 = 0$

$7I_1 + 6I_2 + 4I_3 = 8$ (i)

**Circuit BCDFB**. Starting from point B, we get

$0.4I_2 - 10 - 1.5I_3 + 3(I_1+I_2) = 0$

$30I_1 + 34I_2 - 15I_2 = 100$ (ii)

**Circuit FDEGF**

$+0.3(I_2+I_3) - 6 + 6(I_1+I_2+I_3) + 1.5I_3 = 0$ (iii)

Solving for the three currents, we get

$I_1 = -1.25$ A

$I_2 = 3.54$A

$I_3 = -1.13$ A

The negative signs mean that actual direction of flow of $I_1$ and $I_3$ are opposite to those shown in Fig 3.8 .

6 Use KCL to *find the current supplied by the voltage-controlled current source in Fig 3.9.*

***Solution.*** We will apply KCL to node A. currents coming towards *A* would be taken positive and those going away from it as negative

$-2 - I_1 + 2\upsilon - I_2 = 0$

$2\upsilon = I_1 + I_2 + 2$

Now, $I_1 = \upsilon/3$ and $I_2 = \upsilon/6$

$2\upsilon = \upsilon/3 + \upsilon/6 + 2$

$\upsilon = 4/3$

Hence, value; of current source = $2 \times 4/3 = 8/3$ A

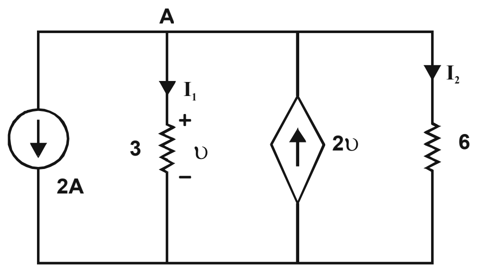



*Fig.3.9*

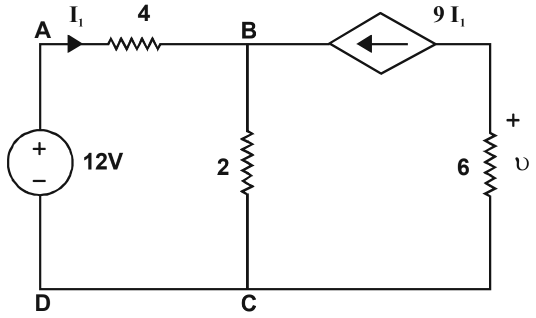



*Fig.3.10*

7. Using KVL, find the value of $\upsilon$ in the circuit of Fig. 3.10

***Solution.*** As seen, in this circuit, the value of the dependent current source depends on the current $I_1$ through the 4Ω resistor. It is obvious that current through 2Ω resistor is 10 $I_1$. Let us apply KVL to the closed circuit ABCDA. Starting from point A in the clockwise direction, we have

$-4I_1 - (10I_1 \times 2) + 12 = 0$

or $I_1 = 0.5$ A

Now, current through 6Ω resistor is $= 9I_1 = 9 \times 0.5 = 4.5$ A.

Hence $\upsilon = 6 \times 4.5 = 27$ V

8. Find the current in each branch of the given network in fig.3.11

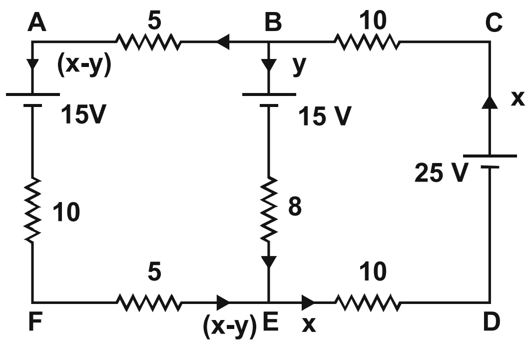



*Fig.3.11*

**Solution.** Starting from point B and applying KVL to circuit BAFEB, we get

$-5(x-y) - 15 - 15(x-y) + 8y + 15 = 0$

$5x = 7y$ (i)

From circuit CBEDC, we get

$-10x - 15 - 8y - 10x + 25 = 0$

$10x + 4y = 5$ (ii)

From (i) and (ii), we get

$x = 7/18$ A;

$y = 5/18$ A

■■■

# CHAPTER : 4
# ELECTROMAGNETIC INDUCTION

**RELATION BETWEEN MAGNETISM AND ELECTRICITY**

It is well known that whenever an electric current flows through a conductor, a magnetic field is immediately brought into existence in the space surrounding the conductor. It can be said that when electrons are in motion, they produce a magnetic field. The converse of this is also true i.e. when a magnetic field embracing a conductor moves relative to the conductor, it produces a flow of electrons in the conductor. This phenomenon whereby an e.m.f. and hence current (i.e. flow of electrons) is induced in any conductor which is cut across or is cut by a magnetic flux is known as electromagnetic induction. The historical background of this phenomenon is this:

After the discovery (by Oersted) that electric current produces a magnetic field, scientists began to search for the converse phenomenon from about 1821 onwards. The problem they put to themselves was how to 'convert' magnetism into electricity. It is recorded that Michael Faraday was in the habit of walking about with magnets in his pockets so as to constantly remind him of the problem. After nine years of continuous research and experimentation, he succeeded in producing electricity by 'converting magnetism'. In 1831, he formulated basic laws underlying the phenomenon of electromagnetic induction (known after his name), upon which is based the operation of most of the commercial apparatus like motors, generators and transformers etc.

**PRODUCTION OF INDUCED E.M.F. AND CURRENT**

In Fig. 4.1 is shown an insulated coil whose terminals are connected to a sensitive galvanometer G. It is placed close to a stationary bar magnet initially at position *AB* (shown dotted). As seen, some flux from the N-pole of the magnet is linked with or threads through the coil but, as yet, there is no deflection of the galvanometer. Now, suppose that the magnet is suddenly brought closer to the coil in position CD (see figure). Then, it is found that there is a jerk or a sudden but a momentary deflection in the galvanometer and that this lasts so long as the magnet is in motion relative to the coil, not otherwise. The deflection is reduced to zero when the magnet becomes again stationary at its new position CD. It should be noted that due to the approach of the magnet, flux linked with the coil is increased.

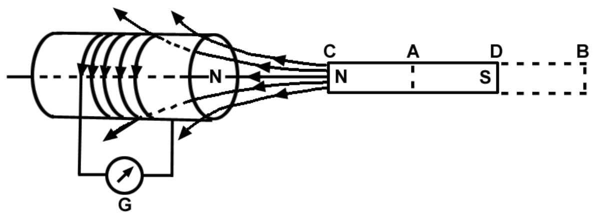



*Fig.4.1*

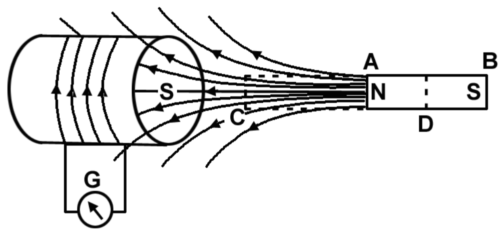



*Fig.4.2*

Next, the magnet is suddenly withdrawn away from the coil as in Fig. 4.2. It is found that again there is a momentary deflection in the galvanometer and it persists so long as the magnet is in motion, not when it becomes stationary. It is important to note that this deflection is in a direction opposite to that of Fig. 4.1. Obviously, due to the withdrawal of the magnet, flux linked with the coil is decreased.

The deflection of the galvanometer indicates the production of e.m.f. in the coil. The only cause of the production can be the sudden approach or withdrawal of the magnet from the coil. It is found that the actual cause of this e.m.f is the change of flux linking with the coil. This e.m.f. exists so long as the change in a stationary conductor. In fact, the same results can be obtained by keeping the bar magnet stationary and moving the coil suddenly away or towards the magnet.

The direction of current set up by the induced e.m.f. is as shown in the two figures given above.

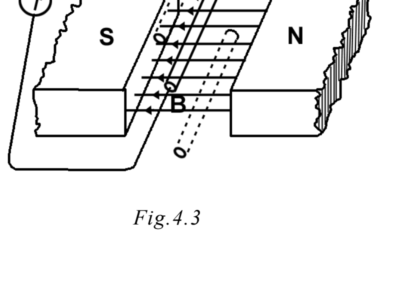



*Fig.4.3*

The production of this electromagnetically induced e.m.f. is further illustrated by considering a conductor AB lying within a magnetic field and connected to a galvanometer as shown in Fig. 4.3. It is found that whenever this conductor is moved up or down, a momentary deflection is produced in the galvanometer. It means that some transient e.m.f. is induced in AB. The magnitude of this induced e.m.f is induced in AB. The magnitude of this induced e.m.f (and hence the amount of deflection in the galvanometer) depends on the quickness of the movement of AB.

From this experiment we conclude that whenever a conductor cuts or shears the magnetic flux, an e.m.f. is always induced in it.

It is also found that if the conductor is moved parallel to the direction of the flux so that it does not cut it, then no e.m.f is produced in it.

**FARADAY'S LAWS OF ELECTROMAGNETIC INDUCTION**

Faraday summed up the above facts into two laws known as Faraday's Laws of Electromagnetic Induction.

**First Law.** It states:-

Whenever the magnetic flux linked with a circuit changes, an e.m.f. is always induced in it.

or

Whenever a conductor cuts magnetic flux, an e.m.f. is induced in that conductor.

**Second Law.** It states :-

The magnitude of the induced e.m.f. is equal to the rate of change of flux-linkages.

**Explanation.** Suppose a coil has N turns and flux through it changes from an initial value of $\phi_1$ webers to the final value of $\phi_2$ webers in time t seconds. Then, remembering that by flux-linkages is meant the product of number of turns and the flux linked with the coil, we have

Initial flux linkages = $N\Phi_1$

Final flux linkages = $N\Phi_2$

$\therefore$ Induced e.m.f. $$e = \frac{N\phi_2 - N\phi_1}{t} \text{ volt}$$

or $$e = N\frac{\phi_2 - \phi_1}{t} \text{ volt}$$

Putting the above expression in its differential form, we get

$$e = \frac{d}{dt}(N\phi) \qquad \text{or} \qquad e = N\frac{d\phi}{dt} \text{ volt}$$

Usually, a minus sign is given to the right-hand side expression to signify the fact that the induced e.m.f sets up current in such a direction that magnetic effect produced by it opposes the very cause producing it (Art. 7-5)

$$\therefore \quad e = -N\frac{d\phi}{dt} \text{ volt}$$

**DIRECTION OF INDUCED E.M.F. AND CURRENT**

There exists a definite relation between the direction of the induced current, the direction of the flux and the direction of motion of the conductor. The direction of the induced current may be found easily by applying either Fleming's Right-hand Rule or Flat-hand rule or Lenz's Law.

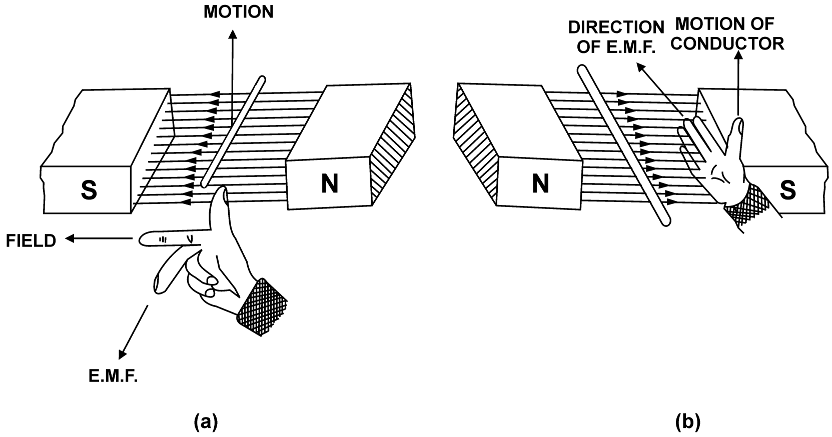



*Fig.4.4*

Fleming's rule (Fig. 4.4) is used where induced e.m.f. is due to flux-cutting (i.e. dynamically induced e.m.f.) and Lenz's when it is due to change by flux-linkages (i.e. statically induced e.m.f)

Fig. 4.5 shows another way of finding the direction of the induced e.m.f. It is known as Right Flat-hand rule. Here, the front side of the hand is held perpendicular to the incident flux with the thumb pointing in the direction of the motion of the conductor. The direction of the fingers gives the direction of the induced e.m.f. and current.

**Lenz's Law**

The direction of the induced current may also be found by this law which was formulated by Lenz in 1835. This law states, in effect, that electromagnetically induced current always flows in such direction that the action of the magnetic field set up by it tends to oppose the very cause which produces it.

This statement will be clarified with reference to Fig. 4.1 and 4.2. It is found that when N-pole of the bar magnet approaches the coil, the induced current set up by induced e.m.f. flows in the anticlockwise direction in the coil as seen from the magnet side. The result is that face of the coil becomes a N-pole and so tends to retard the onward approach of the N-pole of the magnet (like poles repel each other). The mechanical energy spent in overcoming this repulsive force is converted into electrical energy which appears in the coil.

When the magnet is withdrawn as in Fig. 4.2, the induced current flows in the clockwise direction thus making the face of the coil (facing the magnet) a S-pole. Therefore, the N-pole of the magnet has to be withdrawn against this attractive force of the S-pole of coil. Again , the mechanical energy required to overcome this force of attraction is converted into electric energy.

It can be shown that Lenz's law is a direct consequence of Law of Conservation of Energy. Imagine for a moment that when N-pole of the magnet (Fig. 4.1) approaches the coil, induced current flows in such a direction as to make the coil face a S-pole. Then, due to inherent attraction between unlike poles, the magnet would be automatically pulled towards the coil without the expenditure of any mechanical energy. It means that we would be able to create electric energy out of nothing which is denied by the inviolable Law of Conservation of Energy. In fact, to maintain the sanctity of this law, it is imperative for the induced current to flow in such a direction that the magnetic effect produced by it tends to oppose the very cause which produces it. In the present case, it is the relative motion of the magnet with respect to the coil which is the cause of the production of the induced current. Hence, the induced current always flows in such a direction as to tend to oppose this relative motion i.e. the approach or withdrawal of the magnet.

**INDUCED E.M.F.**

Induced e.m.f. can be either (i) dynamically induced or (ii) statically induced. In the first case, usually the field is stationary and conductors cut across it (as in d.c. generators). But in the second case, usually the conductor or the coil remains stationary and flux linked with it is changed by simply increasing or decreasing the current producing this flux (as in transformers)

**DYNAMICALLY INDUCED E.M.F.**

In Fig. 4.5, a conductor A is shown in cross-section lying within a uniform magnetic field of flux density B Wb/m$^2$. The arrow attached to A shows its direction of motion Consider the condition shown in Fig. 4.5 (a) when A cuts across at right angles to the flux. Suppose '$\ell$' is its length lying within the field and let it move a distance dx in time dt. Then area swept by it is = $\ell$ dx. Hence, flux cut = $\ell$ .dx . B webers.

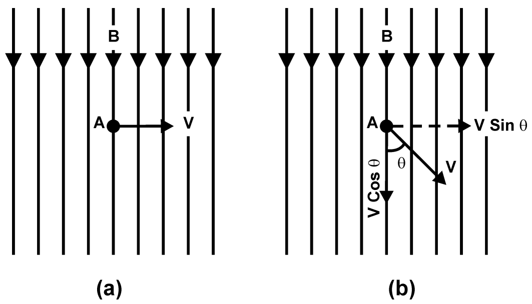



*Fig.4.5*

Change in flux = Bldx weber
Time taken = dt second
hence, according to Faraday's Laws the e.m.f. induced in it (known as dynamically induced e.m.f.) is
= rate of change of flux linkages

$$= \frac{d\phi}{dt} = \frac{Bldx}{dt} = Bl\frac{dx}{dt} = Blv \text{ volt}$$

where $\frac{dx}{dt}$ = conductor velocity $\upsilon$

If the conductor A moves at an angle θ with the direction of flux [Fig. 4.5 (b) ] then the induced e.m.f is

$$e = Bl\upsilon \sin\theta \text{ volt}$$

The direction of the induced e.m.f. is given by Fleming's Right-hand rule or Flat-hand rule.

It should be noted that generators work on the production of dynamically induced e.m.f. in the conductors housed in a revolving armature lying within a strong magnetic field.

**STATICALLY INDUCED E. M. F**

It can be further subdivided into (a) mutually induced e.m.f. and (b) self-induced e.m.f.

**(a) Mutually-induced e.m.f.**

Consider two coils A and B lying close to each other but not touching each other [Fig 4.6 (a)].

Coil A is joined to a battery, a switch and a variable resistance R whereas coil B is connected to a sensitive voltmeter V. When current through A is established by closing the switch, its magnetic field is set up which partly links with or threads through the coil B. As current through A is changed, the flux linked with B is also changed. Hence, mutually induce e.m.f is produced in B whose magnitude is given by Faraday's Laws and direction by Lenz's Law.

If, now, battery is connected to B and the voltmeter across A [Fig. 4.6 (b)], then the situation is reversed and now a change of current in B will produce mutually induced e.m.f. in A.

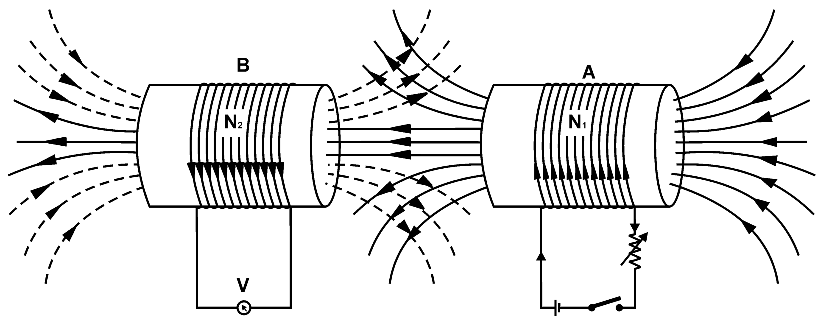



*Fig.4.6 (a)*

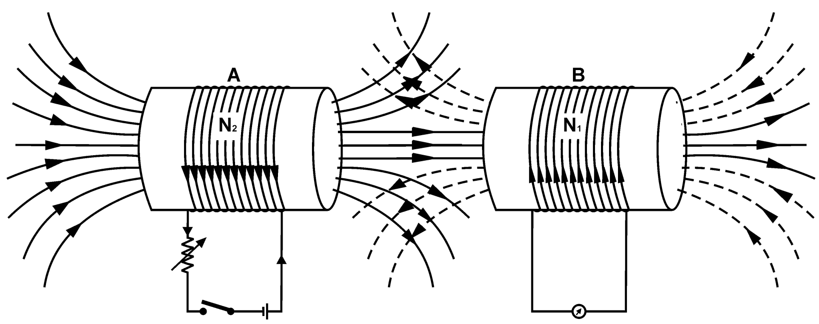



*Fig.4.6 (b)*

It is obvious that in the examples considered above, there is no movement of any conductor, the flux variations being brought about by variations in current strength only. Such an e.m.f. induced in one coil by the influence of the other coil is called (stationary but) mutually induced e.m.f.

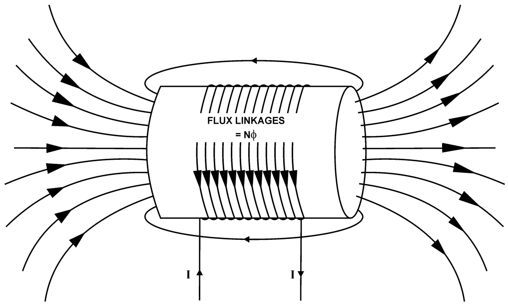



*Fig.4.7*

**(b) Self-induced e.m.f.**

This is the e.m.f. induced in a coil due to the change of its own flux linked with it. If current through the coil (Fig. 4.7) is changed, then the flux linked with its own turns will also change which will produce in it what is called self-induced e.m.f The direction of this induced e.m.f (as given by Lenz's law) would be such as to oppose any change of flux which is, in fact, the very cause of its production. Hence, it is also known as the opposing or counter e.m.f. of self-induction.

**Self-inductance**

Imagine a coil of wire similar to the one shown in Fig. 4.7 connected to a battery through a rheostat. It is found that whenever an effort is made to increase current (and hence flux) through it, it is always opposed by the instantaneous production of counter e.m.f. of self-induction. Energy required to overcome this opposition is supplied by the battery. As will be fully explained later on, this energy is stored in the additional flux produced.

If, now, an effort is made to decrease the current (and hence the flux ), then again it is delayed due to the production of self-induced e.m.f, this time in the opposite direction. This property of the coil due to which it opposes any increase or decrease of current or flux through it, is known as self-inductance. It is quantitatively measured in terms of coefficient of self inductance L. This property is analogous to inertia in a material body. We know by experience that initially it is difficult to set a heavy body into motion, but once in motion. It is equally difficult to stop it. Similarly, in a coil having large self-induction, it is difficult to withdraw it. Hence, self-induction is sometimes analogously called electrical inertia or electromagnetic inertia.

**Coefficient of Self-induction (L)**

The coefficient of self-induction of a coil is defined as :-

"the weber-turns per ampere in the coil"

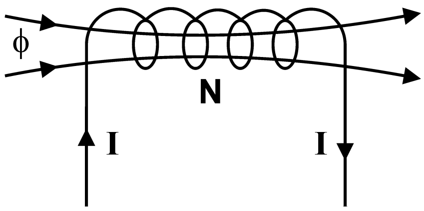



*Fig.4.8*

By 'weber-turns' is meant the product of flux in webers and the number of turns with which the flux is linked. In other words, it is the flux-linkages of the coil.

Consider a solenoid having N turns and carrying a current of I amperes. If the flux produced is $\phi$ webers, then weber-turns are $N\phi$. Hence, weber turns per ampere are $N\phi /I$

By definition, $$L = \frac{N\phi}{I}$$

The unit of self-induction is henry*

If in the above relation,

$N\Phi = 1$ Wb-turn, $I = 1$ ampere, then $L = 1$ henry (H)

Hence, a coil is said to have a self-inductance of one henry if a current of 1 ampere when flowing through it produces flux-linkages of 1 Wb-turn in it.

Therefore, the above relation becomes $L = \dfrac{N\phi}{I}$ henry

**Mutual Inductance**

We have seen that any change of current in coil A is always accompanied by the production of mutually-induced e.m.f in coil B. Mutual inductance may, therefore, be defined as the ability of one coil (or circuit) to produce an e.m.f. in a nearby coil by induction when the current in the first coil changes. This action being reciprocal, the second coil can also induce an e.m.f in the first when current in the second coil changes. This ability of reciprocal induction is measured in terms of the coefficient of mutual induction M.

**Coefficient of Mutual Inductance (M)**

Let there be two magnetically-coupled coils having $N_1$ and $N_2$ turns respectively (Fig. 4.9)

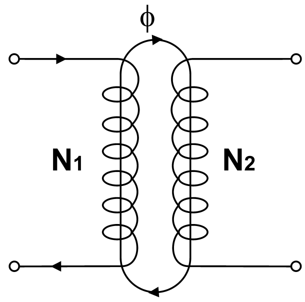



***Fig. 4.9***

Coefficient of mutual inductance between the two coils is defined as the weber-turns in one coil due to one ampere current in the other.

Let a current of $I_1$ ampere when flowing in the first coil produce a flux $\phi$ webers in it. It is supposed that whole of this flux links with the turns of the second coil. * Then, flux-linkages i.e. weber-turns in the second coil for unit current in the first coil are $N_2\Phi_1 / I_1$. Hence, by definition

$$M = N_2 \Phi_1 / I_1$$

If weber-turns in second coil due to one ampere current in the first coil i.e. $N_2 \Phi_1 / I_1 = 1$ then, as seen from above, M=1 H.

Hence, two coils are said to have a mutual inductance of 1 henry if one ampere current when flowing in one coil produces flux-linkages of one Wb-turn in the other.

**Coefficient of Coupling**

Consider two magnetically-coupled coils A and B having $N_1$ and $N_2$ turns respectively. Their individual coefficients of self-induction are,

$$L_1 = [N_1^2 / (l/\mu_o\mu_r A)] \quad \text{and}$$

$$L_2 = [N_2^2 / (l/\mu_o\mu_r A)]$$

The flux $\Phi_1$ produced in A due to a current $I_1$ ampere is

$$\Phi_1 = [N_1 I_1 / (l/\mu_o\mu_r A)]$$

Suppose a fraction $k_1$ of this flux i.e. $k_1 \phi_1$ is linked with coil B.

Then $M = (k_1 \phi_1 \text{ x } N_2)/I_1$ where $k_1 < 1$.

Substituting the value of $\phi_1$ ,

we have, $M = [k_1 \text{ x } \{N_1 N_2 / (I/\mu_o\mu_r A)\}$ (i)

Similarly, the flux $\phi_2$ produced in B due to $I_2$ ampere in it is

$$\phi_2 = \{N_2 I_2 / (l/\mu_o\mu_r A)$$

Suppose a fraction $k_2$ of this flux i.e. $k_2 \phi_2$ is linked with A

Then $M = \{k_2 \phi_2 \text{ x } N_1)/I_2\} = k_2 \{(N_1 N_2)/(I/\mu_o\mu_r A)\}$

Multiplying Eq. (i) and (ii), we get

$$M^2 = [k_1 k_2 \{(N_1)/(l/\mu_o\mu_r A\}] \text{ x } [(N_2)/(I/\mu_o\mu_r A)] \quad \text{or} \quad M^2 = k_1 k_2 L_1 L_2$$

Putting $\sqrt{k_1 k_2} = k$ , we have $M^2 = k^2 L_1 L_2$ or $k = M/\sqrt{L_1 L_2}$

The constant k is called the coefficient of coupling and may be defined as the ratio of mutual inductance actually present between the two coils to the maximum possible value. If the flux due to one coil completely links with the other, then value of k is unity. If the flux of one coil does not at all link with the other, then k = 0 in the first case, when k = 1, coils are said to be tightly coupled and when k= 0, the coils are magnetically isolated from each other.

**Inductances in Series**

i) Let the two coils be so joined in series that their fluxes (or m.m.fs) are additive i.e. in the same direction (Fig.10)

Let $M$ = Coefficient of mutual inductance

$L_1$ = Coefficient of self-inductance of 1st coil

$L_2$ = Coefficient of self-inductance of 2nd coil

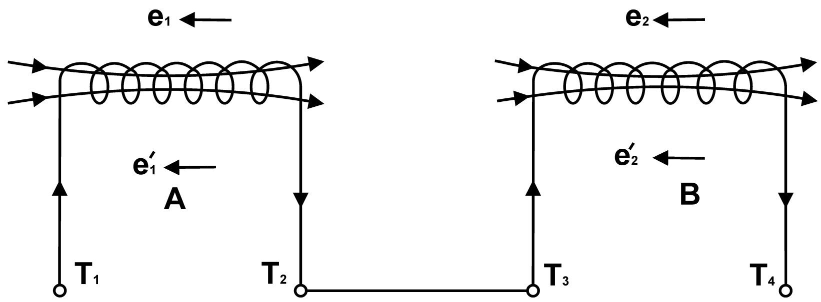



*Fig. 4.10*

Then, self-induced e.m.f. in A is

$$= e_1 = -L_1.\frac{dI}{dt}$$

Mutually-induced e.m.f. in A due to change of current in B is

$$= e_1^{'} = -M.\frac{dI}{dt}$$

Self-induced e.m.f. in B is

$$= e_2 = -L_2.\frac{dI}{dt}$$

Mutually-induced e.m.f. in B due to change of current in A is

$$= e_2^{'} = -M.\frac{dI}{dt}$$

(All have -ve sign, because both self and mutually-induced e.m.fs. are in opposition to the applied e.m.f.)

Total induced e.m.f. in the combination

$$= -\frac{dI}{dt}(L_1 + L_2 + 2M) \qquad \text{(i)}$$

If L is the equivalent inductance, then total induced e.m.f. in that single coil would have been

$$= -L\frac{dI}{dt}$$

Equating (i) and (ii) above, we have $L = L_1 + L_2 + 2M$

ii) When the coils are so joined that their fluxes are in opposite directions (Fig. 11).

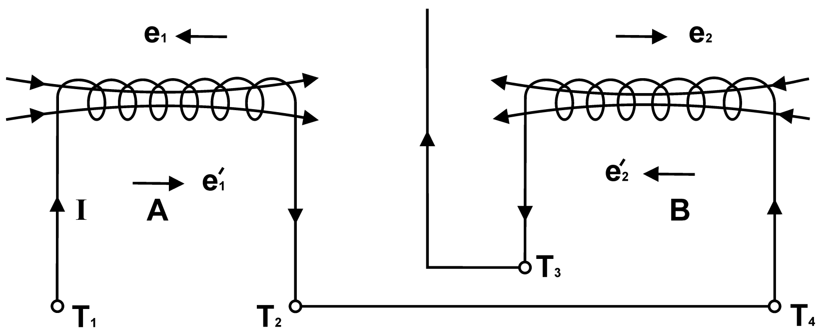



*Fig. 4.11*

As before $e_1 = -L_1 \dfrac{dI}{dt}$

$$e_1^{'} = +M\frac{dI}{dt} \qquad \text{(mark this direction)}$$

$$e_2 = -L_2 \frac{dI}{dt} \quad \text{and} \quad e_2^{'} = +M\frac{dI}{dt}$$

Total induced e.m.f.

$$= -\frac{dI}{dt}\left(L_1 + L_2 - 2M\right)$$

$\therefore$ Equivalent inductance

$$L = L_1 + L_2 - 2M$$

In general, we have

$L = L_1 + L_2 + 2M$ ....if m.m.fs are additive

$\therefore$ $L = L_1 + L_2 - 2M$ .... if m.m.fs. are subtractive

**Inductances in parallel**

In Fig. 4.12, two inductances of values $L_1$ and $L_2$ henry are connected in parallel. Let the coefficient of mutual inductance between the two be M. Let i be the main supply current and $i_1$ and $i_2$ be the branch currents.

Obviously, $i = i_1 + i_2$

$(di/dt) = (di_1/dt) + (di_2/dt)$ (i)

In each coil, both self and mutually induced e.m.fs are produced. Since the coils are in parallel, these e.m.fs are equal. For a case when self-induced e.m.f. assists the mutually-induced e.m.f., we get.

$e = [L_1 (di_1/dt) + M (di_2/dt)] = [L_2 (di_2/dt) + M (di_1/dt)]$

$[L_1 (di_1/dt) + M (di_2/dt)] = [L_2 (di_2/dt) + M (di_1/dt)]$

$[(di_1/dt) (L_1 - M)] = [(di_2/dt)(L_2 - M)]$ $\quad\therefore$ $(di_1/dt) = [(L_2 - M)/ (L_1 - M)] (di_2/dt)$ (ii)

Hence, (i) above becomes $di/dt = [\{(L_2 - M) / (L_1 - M)\} + 1] di_2/dt$ (iii)

If L is the equivalent inductance, then

$e = L (di/dt)$ = induced e.m.f in the parallel combination

= Induced e.m.f in any one coil = $[L_1 (di_1/dt) + M (di_2/dt)]$

$(di_3/ dt) = (I/L) [L_1 (di_1/dt)] + [M (di_2/dt)]$ (iv)

Substituting the value of $di_1dt$ from (ii) in (iv), we get

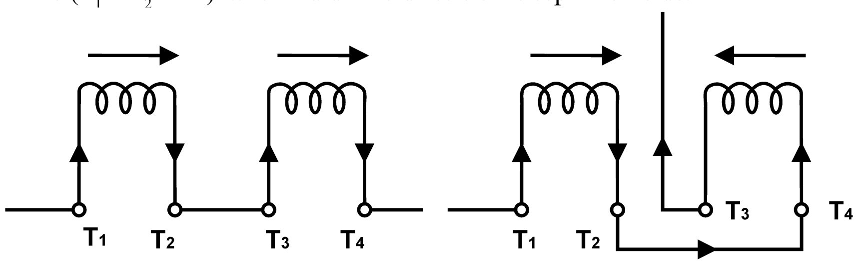
$di_3/dt = (I / L) [ L_1 \{(L_2 - M) / (L_1 - M)\} + M] (di_2/dt)$

Hence, equating (iii) to (iv), we have

$(L_2 - M) / (L_1 - M) + 1 = I /L [ L_1 \{(L_2 - M) / (L_1 - M)\} + M]$

$(L_1 + L_2 - 2M) / ( L_1 - M) = (I/L) (L_1 L_2 - M_2) / ( L_1 - M)$

$L = (L_1 L_2 - M^{2)}/ (L_1 + L_2 - 2M)$ when mutual field assists the separate fields.




**(a)** **(b)**

*Fig. 4.12*

Similarly $L = (L_1 L_2 - M^{2)} / (L_1 + L_2 + 2M)$ when the two fields oppose each other.

**ELECTROMAGNETISM**

In 1819, the Danish physicist, Hans Christian Oersted, discovered that the needle of a compass brought near a current-carrying conductor would be deflected. When the current flow stopped, the compass needle returned to its original position. This important discovery demonstrated a relationship between electricity and magnetism that led to the electromagnet and to many of the inventions on which modern industry is based.

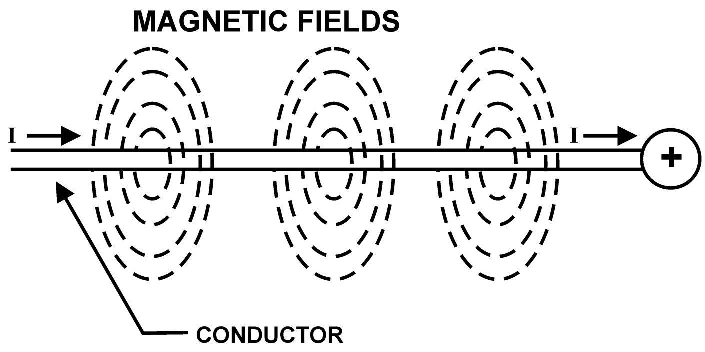



*Fig.4.13, Magnetic held formed around a conductor in which current is flowing.*

Oersted discovered that the magnetic field had no connection with the conductor in which the electrons were flowing, because the conductor was made of nonmagnetic copper. The magnetic field around the conductor was created by the electrons moving through the wire. Since a magnetic field accompanies a charged particle, the greater the current flow the greater the magnetic field. Figure 4.13 illustrates the magnetic field around a current- carrying wire. A series of concentric circles around the conductor represent the field, which, if all the lines were shown, would appear more as a continuous cylinder of such circles around the conductor.

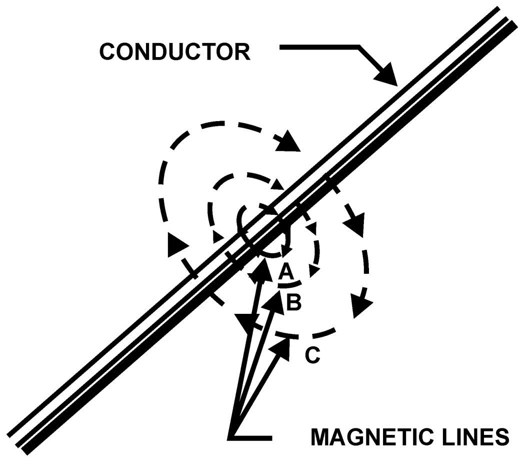



*Fig.4.14, Expansion of magnetic field as current increases.*

As long as current flows in the conductor, the lines of force remain around it, as shown in figure 4.14. If a small current flows through the conductor, there will be a line of force extending out to circle A. If the current flow is increased, the line of force will increase in size to circle B, and a further increase in current will expand it to circle C. As the original line (circle) of force expands from circle A to B, a new line of force will appear at circle A. As the current flow increases, the number of circles of force increases, expanding the outer circles farther from the surface of the current-carrying conductor.

If the current flow is a steady nonvarying direct current, the magnetic field remains stationary. When the current stops, the magnetic field collapses and the magnetism around the conductor disappears.

A compass needle is used to demonstrate the direction of the magnetic field around a current carrying conductor. A of figure 4.15 shows a compass needle positioned at right angles to, and approximately one inch from, a current-carrying conductor. If no current were flowing, the north -seeking end of the compass needle would point toward the earth's magnetic pole. When current flows, the needle lines itself up at right angles to a radius drawn from the conductor. Since the compass needle is a small magnet, with lines of force extending from south to north inside the metal, it will turn until the direction of these lines agrees with the direction of the lines of force around the conductor. As the compass needle is moved around the conductor, it will maintain itself in a position at right angles to the conductor, indicating that the magnetic field around a current-carrying conductor is circular. As shown in B of figure 4.15, when the direction of current flow through the conductor is reversed, the compass needle will point in the opposite direction, indicating the magnetic field has reversed its direction.

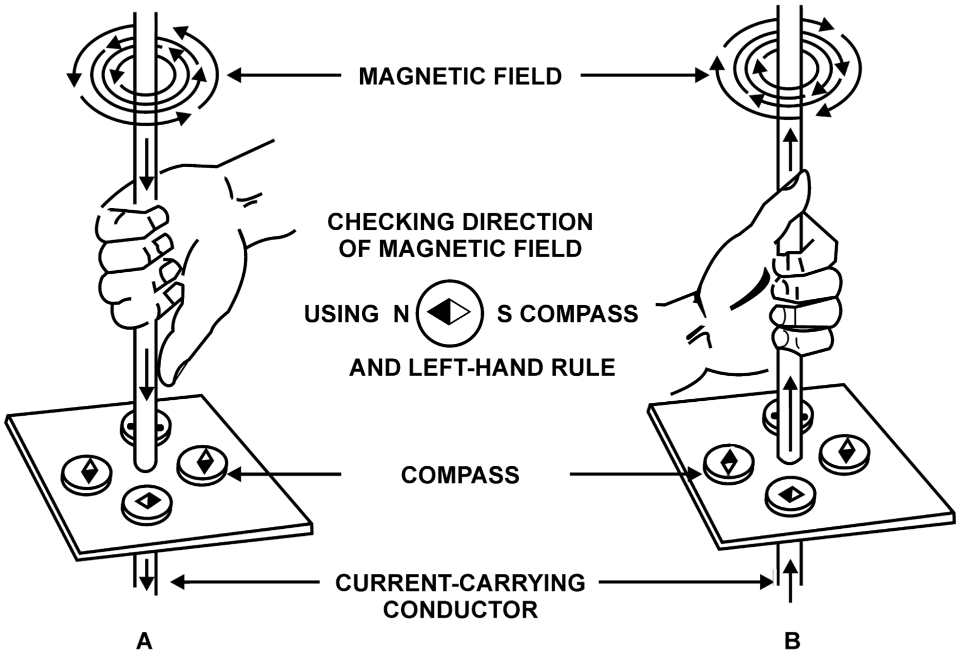



*Fig.4.15, Expansion of magnetic field as current increases.*

A method used to determine the direction of the lines of force when the direction of the current flow is known as shown in figure 4.16. If the conductor is grasped in the left hand, with the thumb pointing in the direction of current flow, the fingers will be wrapped around the conductor in the same direction as the lines of the magnetic field. This is called the left-hand rule.

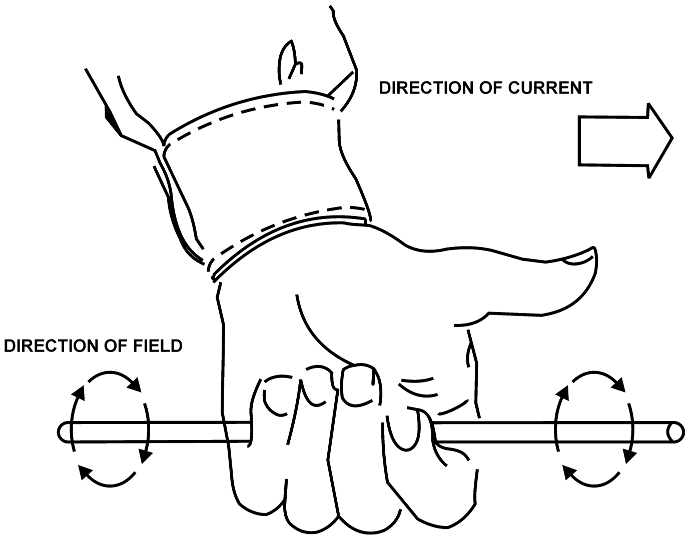



*Fig.4.16, Left-hand rule.*

Although it has been stated that the lines of force have direction, this should not be construed to mean that the lines have motion in a circular direction around the conductor. Although the lines of force tend to act in a clockwise or counter-clockwise direction they are not revolving around the conductor.

Since current flows negative to positive, many illustrations indicate current direction with a dot symbol on the end of the conductor when the current is flowing toward and a cross sign when the current is flowing away from the observer. This is illustrated in figure 4.17.

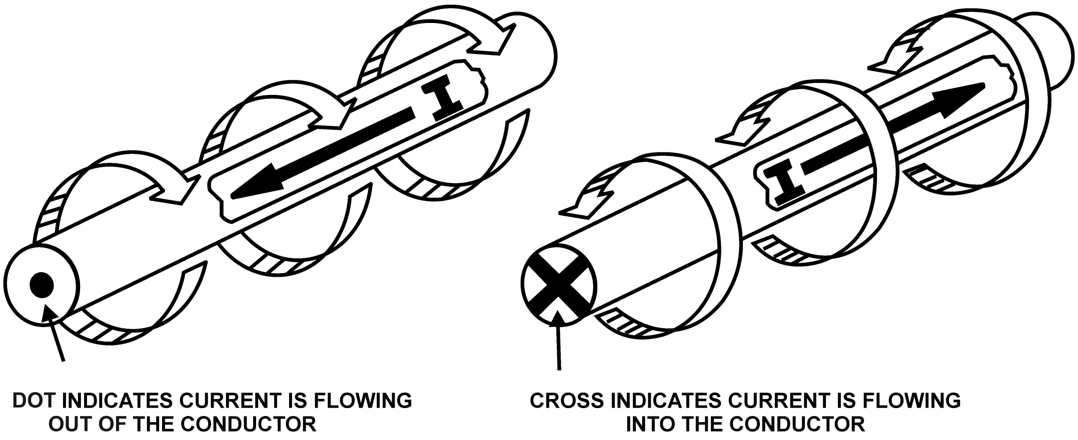



*Fig.4.17, Direction of current flow in a conductor.*

When a wire is bent into a loop and an electric current flows through it, the left-hand rule remains valid, as shown in figure 4.18.

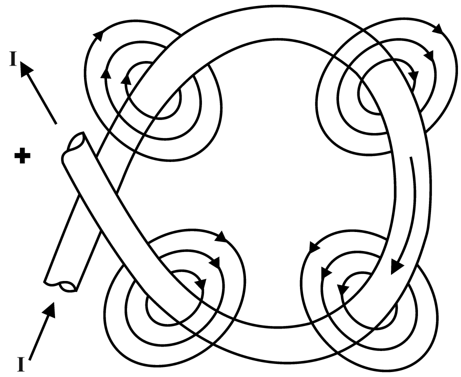



*Fig. 4.18, Magnetic field around a looped conductor.*

If the wire is coiled into two loops, many of the lines of force become large enough to include both loops. Lines of force go through the loops in the same direction, circle around the outside of the two coils, and come in at the opposite end. (See figure 4.19).

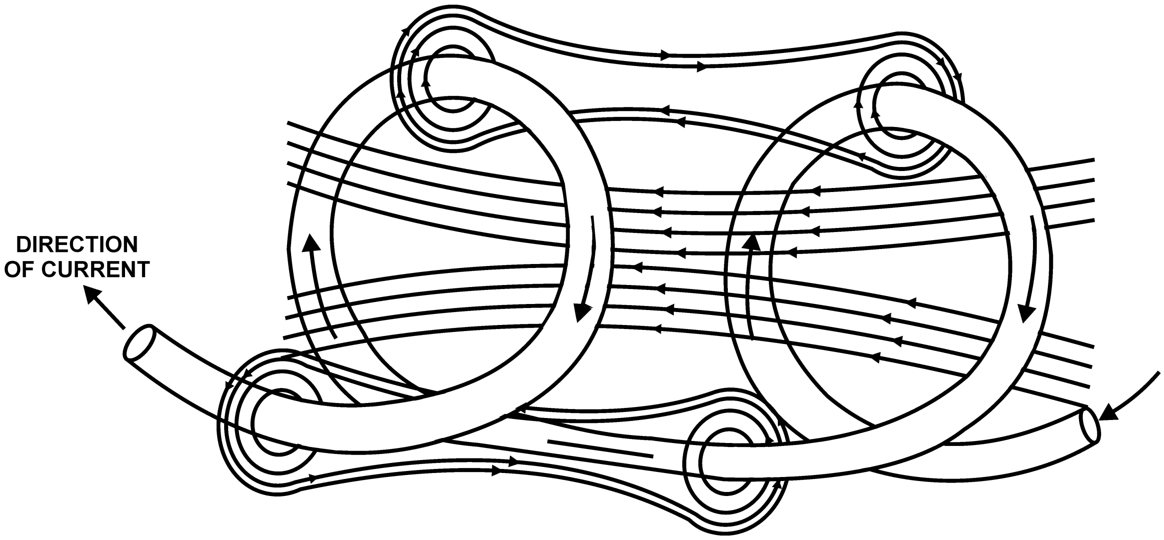



*Fig.4.19, Magnetic field around a conductor with two loops.*

When a wire contains many such loops, it is called a coil. The lines of force form a pattern through all the loops, causing a high concentration of flux lines through the centre of the coil. (See figure 4.20).

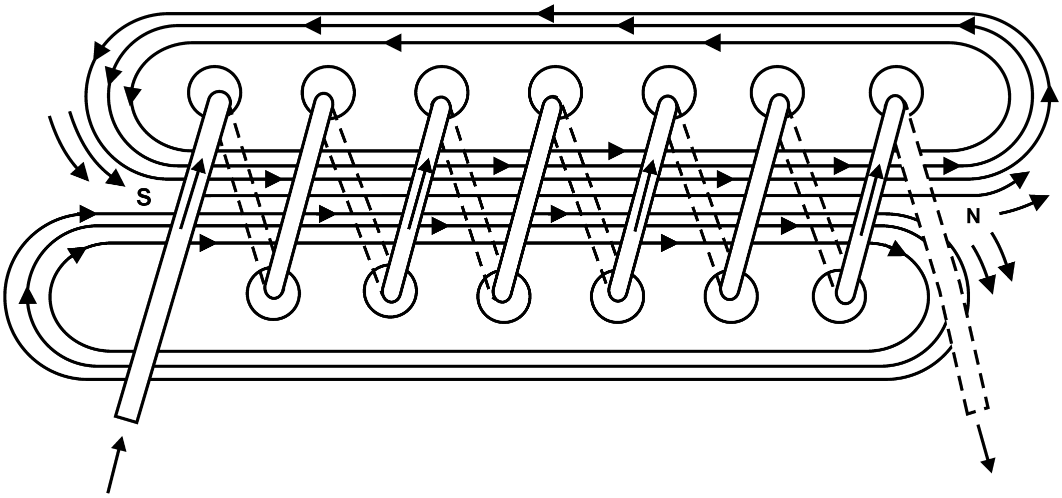



***Fig.4.20, Magnetic field of a coil.***

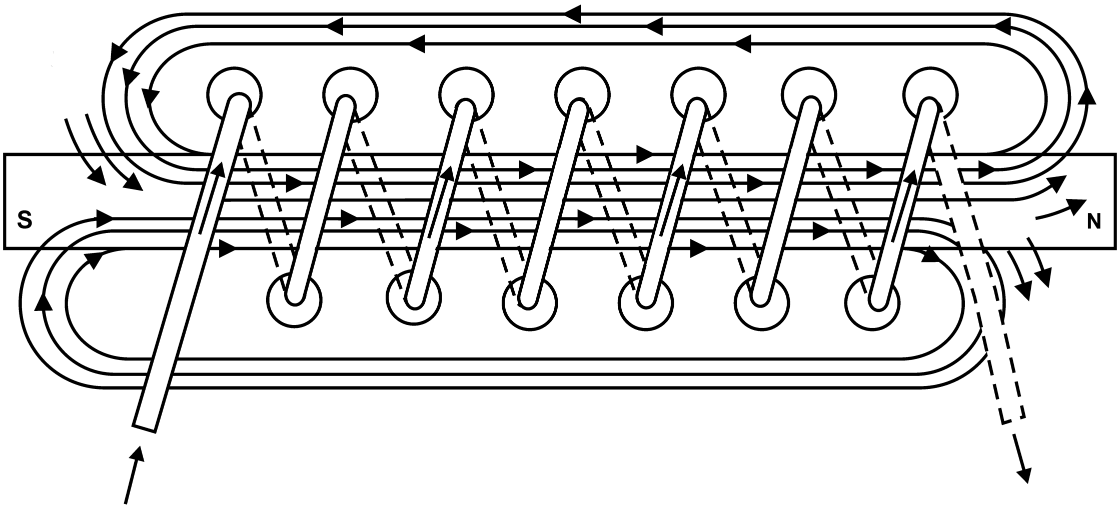



***Fig.4.21, Electromagnet.***

In a coil made from loops of a conductor, many of the lines of force are dissipated between the loops of the coil. By placing a soft iron bar inside the coil, the lines of force will be concentrated in the center of the coil, since soft iron has a greater permeability than air. (See figure 4.21). This combination of an iron core in a coil of wire loops, or turns, is called an electromagnet, since the poles (ends) of the coil possess the characteristics of a bar magnet.

The addition of the soft iron core does two things for the current-carrying coil. First, the magnetic flux is increased, and second, the flux lines are more highly concentrated.

When direct current flows through the coil, the core will become magnetized with the same polarity (location of north and south poles) as the coil would have without the core. If the current is reversed, the polarity will also be reverse.

The polarity of the electromagnet is determined by the left-hand rule in the same manner as the polarity of the coil without the core was determined. If the coil is grasped in the left hand in such a manner that the fingers curve around the coil in the direction of electron flow (minus to plus), the thumb will point in the direction of the north pole. (See figure 4.22).

The strength of the magnetic field of the electromagnet can be increased by either increasing the flow of current or the number of loops in the wire. Doubling the current flow approximately doubles the strength of the field, and in a similar manner, doubling the number of loops approximately doubles magnetic field strength. Finally, the type metal in the core is a factor in the field strength of the electromagnet.

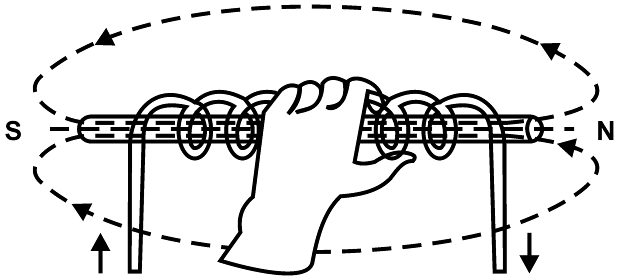



*Fig.4.22, Left-hand rule applied to a coil.*

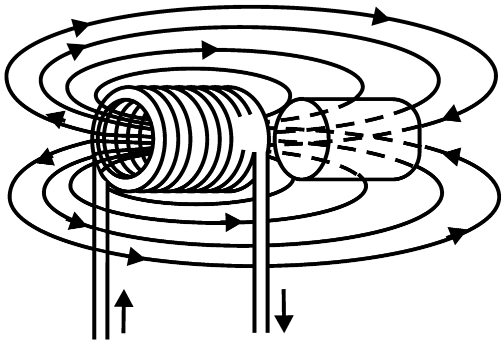


*Fig.4.23, Solenoid with iron core.*

A soft-iron bar is attracted to either pole of a permanent magnet and, likewise, is attracted by a current-carrying coil. As shown in figure 4.23, the lines of force extend through the soft iron, magnetizing it by induction and pulling the iron bar toward the coil. If the bar is free to move, it will be drawn into the coil to a position near the center where the field is strongest.

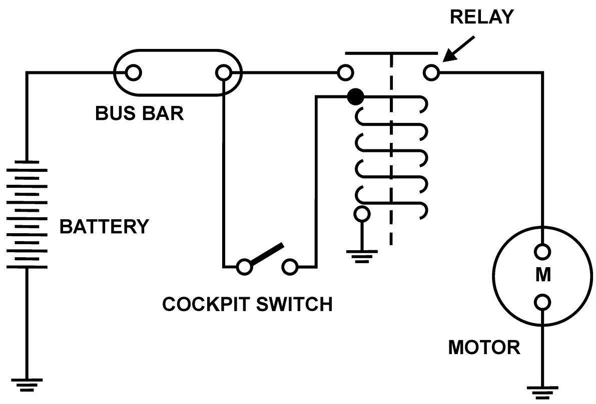



*Fig.4.24, Use of a solenoid in a circuit.*

Electromagnets are used in electrical instruments, motors, generators, relays, and other devices. Some electromagnetic devices operate on the principle that an iron core held away from the center of a coil will be rapidly pulled into a center position when the coil is energized. This principle is used in the solenoid, also called solenoid switch or relay, in which the iron core is spring-loaded off center and moves to complete a circuit when the coil is energized.

The application of the solenoid is shown in figure 4.24, where it is a solenoid relay. When the cockpit switch is closed, the energized coil pulls the core switch down, which completes the circuit to the motor. Since this solenoid relay operates on low current, it eliminates high-amperage wiring in the cockpit of the aircraft.

The solenoid-and-plunger type of magnet in various forms is used extensively to open circuit breakers automatically, when the load current becomes excessive, and the operate valves, magnetic brakes, and many other devices. The armature-type of electromagnet also has extensive applications. For this type of magnet, the coil is wound on and insulated from the iron core; the core is not movable. When current flows through the coil, the iron core becomes magnetized and causes a pivoted soft-iron armature located near the electromagnet to be attracted to it. These magnets are used in doorbells, relays, circuit breakers, telephone receivers, and many other devices.

■■■

# CHAPTER : 5
# DIRECT CURRENT GENERATOR

**DC Generator Principle**
An electrical generator is a machine which converts mechanical energy (or power) into electrical energy (or power).

The energy conversion is based on the principle of the production of dynamically (or motionally) induced e.m.f. As seen from Fig. 5.1., whenever a conductor cuts magnetic flux, dynamically induced e.m.f. is produced in it according to Faraday's Laws of Electromagnetic Induction. This e.m.f. causes a current to flow if the conductor circuit is closed.

Hence, two basic essential parts of an electrical generator are (i) a magnetic field and (ii) a conductor or conductors which can so move as to cut the flux.

**Simple Loop Generator**
**Construction**
In Fig.5.1 is shown a single-turn rectangular copper coil *ABCD* rotating about its own axis in a magnetic field provided by either permanent magnets or electromagnets. The two ends of the coil are joined to two slip-rings *'a' and 'b'* which

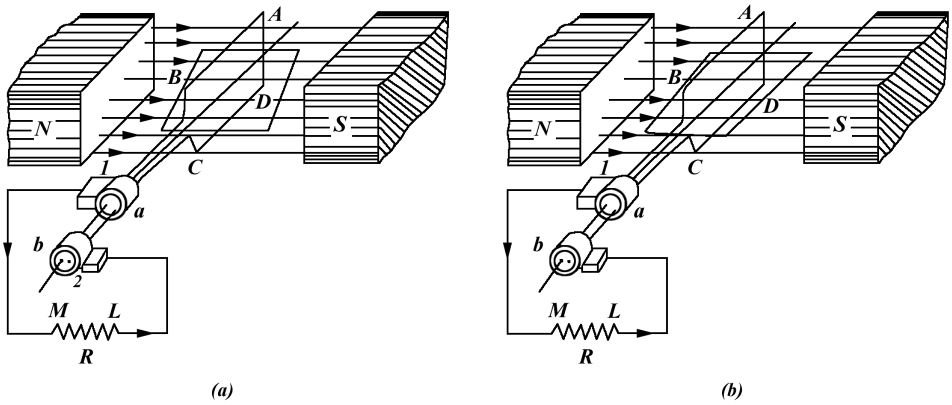



*Fig.5.1*

are insulated from each other and from the central shaft. Two collecting brushes (of carbon or copper) press against the slip-rings. Their function is to collect the current induced in the coil and to convey it to the external load resistance *R*.

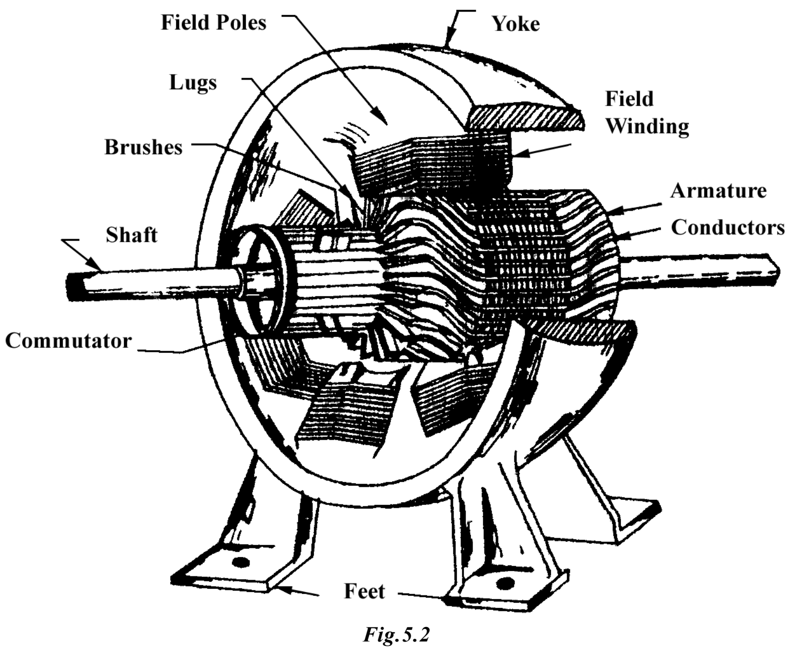



*Fig.5.2*

**Practical Generator**

The simple loop generator has been considered in detail merely to bring out the basic principle underlying construction and working of an actual generator illustrated in Fig. 5.2 which consists of the following essential parts :

1. Magnetic Frame or Yoke
2. Pole-Cores and Pole-Shoes
3. Pole Coils
4. Armature Core
5. Armature Windings or Conductors
6. Commutator
7. Brushes and Bearings

Of these, the yoke, the pole cores, the armature core, the air gaps between the poles and the armature core form the magnetic circuit whereas the rest form the electrical circuit.

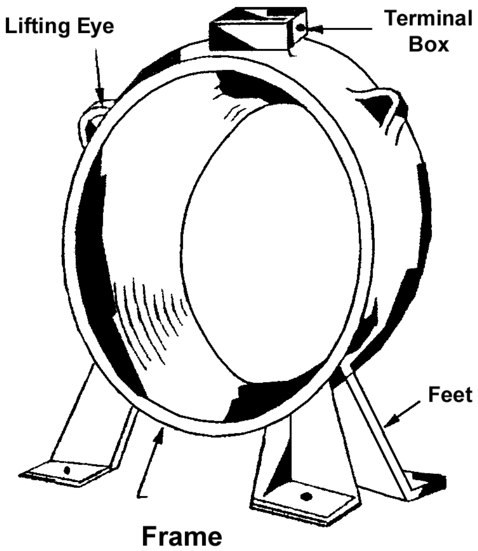


*Fig.5.3*

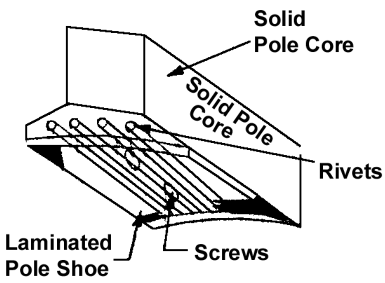


*Fig.5.4*

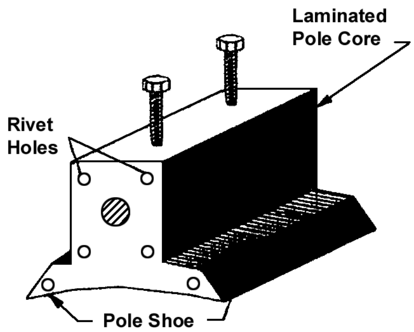


*Fig.5.5*

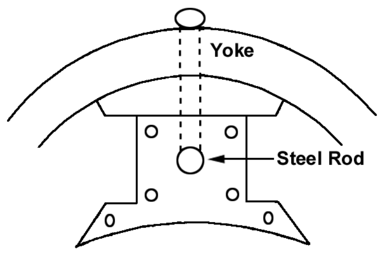


*Fig.5.6*

**Yoke**

The outer frame or yoke serves double purpose :

i. It provides mechanical support for the poles and acts as a protecting cover for the whole machine and
ii. It carries the magnetic flux produced by the poles.

In small generators where cheapness rather than weight is the main consideration, yokes are made of cast iron. But for large machines usually cast steel or rolled steel is employed. The modern process of forming the yoke consists of rolling a steel slab round a cylindrical mandrel and then welding it at the bottom. The feet and the terminal box etc. are welded to the frame afterwards. Such yokes possess sufficient mechanical strength and have high permeability.

**Pole Cores and Pole Shoes**

The field magnets consist of pole cores and pole shoes. The pole shoes serve two purposes (i) they spread out the flux in the air gap and also, being of larger cross-section, reduce the reluctance of the magnetic path (ii) they support the exciting coils (or field coils) as shown in Fig. 5.8.

There are two main types of pole construction.

a. The pole core itself may be a solid piece made out of either cast iron or cast steel but the pole shoe is laminated and is fastened to the pole face by means of counter sunk screws.
b. In modern design, the complete pole cores and pole shoes are built of thin laminations of annealed steel which are revetted together under hydraulic pressure (Fig. 5.5). The thickness of laminations varies from 1 mm to 0.25 mm. The laminated poles may be secured to the yoke in any of the following two ways :

i. Either the pole is secured to the yoke by means of screws bolted through the yoke and into the pole body or
ii. The holding screws are bolted into a steel bar which passes through the pole across the plane of laminations (Fig. 5.6).

**Pole Coils**

The field coils or pole coils, which consist of copper wire or strip, are former-wound for the correct dimension (Fig. 5.7). Then, the former is removed and wound coil is put into place over the core as shown in Fig.5.8.

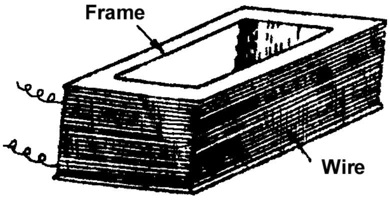



*Fig.5.7*

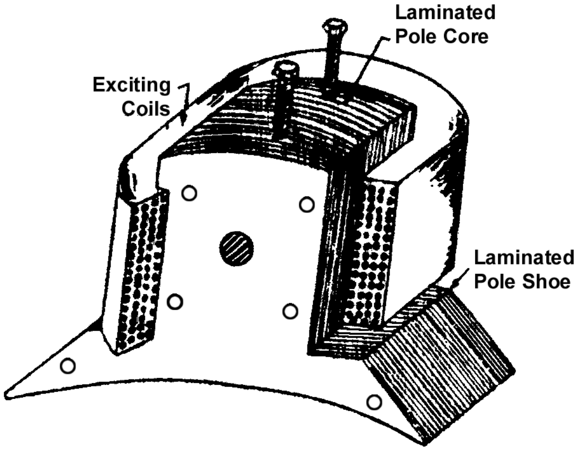



*Fig. 5.8*

When current is passed through these coils, they electromagnetise the poles which produce the necessary flux that is cut by revolving armature conductors.

**Armature Core**

It houses the armature conductors or coils and cause them to rotate and hence cut the magnetic flux of the field magnets. In addition to this, its most important function is to provide a path of very low reluctance to the flux through the armature from a N-pole to a S-pole.

It is cylindrical or drum-shaped and is built up of usually circular sheet steel discs or laminations approximately 0.5 mm thick (Fig. 5.9). It is keyed to the shaft.

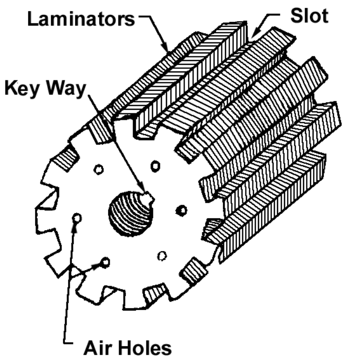



*Fig. 5.9*

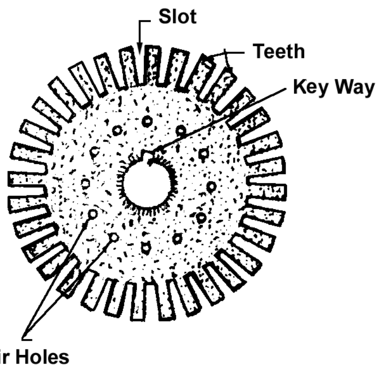



*Fig. 5.10*

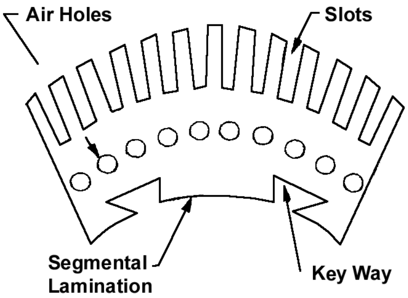



*Fig. 5.11*

The slots are either die-cut or punched on the outer periphery of the disc and the keyway is located on the inner diameter as shown. In small machines, the armature stampings are keyed directly to the shaft. Usually, these laminations are perforated for air ducts which permits axial flow of air through the armature for cooling purposes. Such ventilating channels are clearly visible in the laminations shown in Fig. 5.10 and Fig. 5.11.

Up to armature diameters of about one meter, the circular stampings are cut out in one piece as shown in Fig. 5.10. But above this size, these circles, especially of such thin sections, are difficult to handle because they tend to distort and become wavy when assembled together. Hence, the circular laminations, instead of being cut out in one piece, are cut in a number of suitable sections or segments which form part of a complete ring (Fig. 5.11).

A complete circular lamination is made up of four or six or even eight segmental laminations. Usually, two keyways are notched in each segment and are dove-tailed or wedge-shaped to make the laminations self-locking in position.

The purpose of using laminations is to reduce the loss due to eddy currents. Thinner the laminations, greater is the resistance offered to the induced e.m.f., smaller the current and hence lesser the $I^2R$ loss in the core.

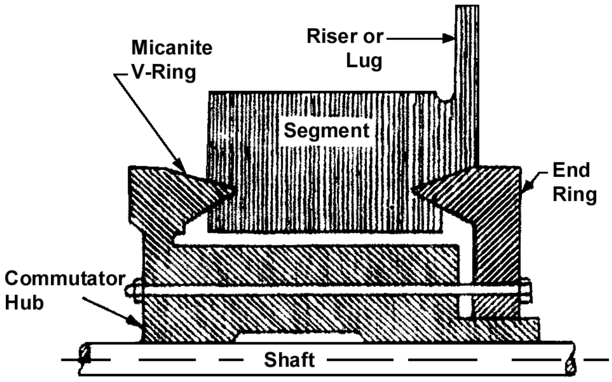



*Fig. 5.12*

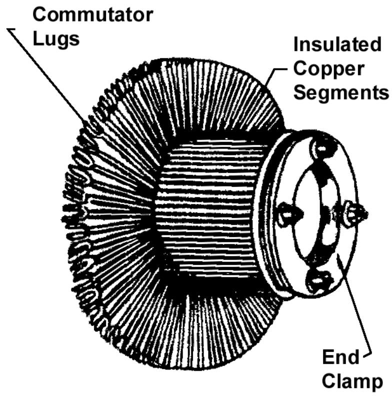



*Fig.5.13*

**Commutator**

The function of the commutator is to facilitate collection of current from the armature conductors. As shown in art. 5.13, it rectifies i.e. converts the alternating current induced in the armature conductors into unidirectional current in the external load circuit. It is cylindrical in structure and is built up of wedge-shaped segments of high-conductivity hard-drawn or drop forged copper. These segments are insulated from each other by thin layers of mica. The number of segments is equal to the number of armature coils. Each commutator segment is connected to the armature conductor by means of a copper lug or strip (or riser). To prevent them from flying out under the action of centrifugal forces, the segments have V-grooves, these grooves being insulated by conical micanite rings. A sectional view of commutator is shown in Fig. 5.12. whose general appearance when completed is shown in Fig. 5.13.

**Armature Windings**

The armature windings are usually former-wound. These are first wound in the form of flat rectangular coils and are then pulled into their proper shape in a coil puller. Various conductors of the coils are insulated from each other. The conductors are placed in the armature slots which are lined with tough insulating material. This slot insulation is folded over above the armature conductors placed in the slot and is secured in place by special hard wooden or fibre wedges.

Now, we will discuss the winding of an actual armature. But before doing this, the meaning of the following terms used in connection with armature winding should be clearly kept in mind.

**Pole-pitch**

It may be variously defined as :

i. The periphery of the armature divided by the number of poles of the generator i.e. the distance between two adjacent poles.
ii. It is equal to the number of armature conductors (or armature slots) per pole. If there are 48 conductors and 4 poles, the pole pitch is 48/4 = 12.

**Conductor**

The length of a wire lying in the magnetic field and in which an e.m.f. is induced, is called a conductor (or inductor) as, for example, length AB or CD in Fig. 5.14.

**Coil and Winding Element**

With reference to Fig. 5.14, the two conductors *AB* and *CD* along with their end connections constitute one coil of the armature winding. The coil may be single-turn coil (Fig. 5.14) or multiturn coil (Fig. 5.15). A single-turn coil will have two conductors. But a multi-turn coil may have many conductors per coil side. In Fig. 5.15, for example, each coil side has 3 conductors. The group of wires or conductors constituting a coil side of a multi-turn coil is wrapped with a tape as a unit (Fig. 5.16) and is placed in the armature slot. It may be noted that since the beginning and the end of each coil must be connected to a commutator bar, there are as many commutator bars as coils for both the lap and wave windings.

The side of a coil (1-turn or multiturn) is called a winding element. Obviously, the number of winding elements is twice the number of coils.

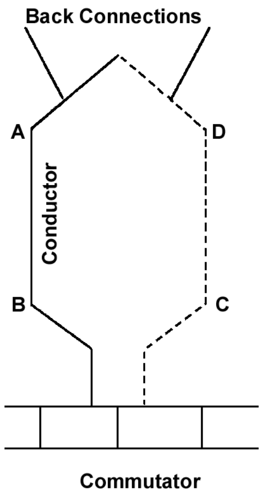



*Fig.5.14*

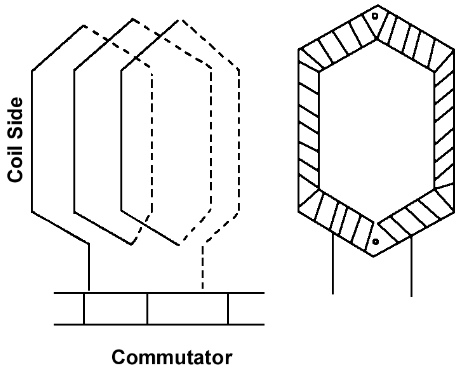



*Fig. 5.15*

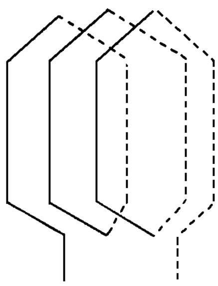


*Fig. 5.16*

**Coil-span or Coil-pitch ($Y_S$)**
It is the distance, measured in terms of armature slots (or armature conductors) between two sides of a coil. It is, in fact, the periphery of the armature spanned by the two sides of the coil.

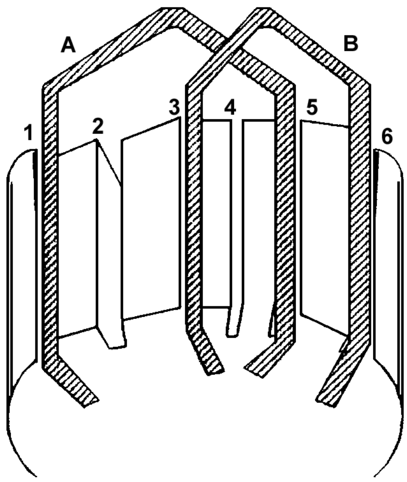



*Fig.5.17*

If the pole span or coil pitch is equal to the pole pitch (as in the case of coil A in Fig. 5.17 where pole-pitch of 4 has been assumed), then winding is called *full-pitched.* It means that coil span is 180 electrical degrees. In this case, the coil sides lie under opposite poles, hence the induced e.m.fs in them are additive. Therefore, maximum e.m.f. is induced in the coil as a whole, it being the sum of the e.m.fs induced in the two coil sides. For example, if there are 36 slots and 4 poles, then coil span is 36/4 = 9 slots. If number of slots is 35, then $Y_S = 35/4 = 8$ because it is customary to drop fractions.

If the coil span is less than the pole pitch (as in coil B where coil pitch is 3/4th of the pole pitch), then the winding is fractional-pitched. In this case, there is a phase difference between the e.m.fs. in the two sides of the coil. Hence,

the total e.m.f. round the coil which is the vector sum of e.m.fs. in the two coil sides, is less in this case as compared to that in the first case.

**Pitch of a Winding (Y)**
In general, it may be defined as the distance round the armature between two successive conductors which are directly connected together. Or, it is the distance between the beginnings of two consecutive turns.

$$Y = Y_B - Y_F \quad \text{........for lap winding}$$
$$= Y_B + Y_F \quad \text{........for wave winding}$$

In practice, coil-pitches as low as eight-tenths of a pole pitch are employed without much serious reduction in the e.m.f. Fractional-pitched windings are purposely used to effect substantial saving in the copper of the end connections and for improving commutation.

**Back Pitch ($Y_B$)**

The distance, measured in terms of the armature conductors, which a coil advances on the back of the armature is called back pitch and is denoted by $Y_B$.

As seen from Fig. 5.21, element 1 is connected on the back of the armature to element 8. Hence, $Y_B = (8 - 1) = 7$.

**Front Pitch ($Y_F$)**

The number of armature conductors or elements spanned by a coil on the front (or commutator end of an armature) is called the front pitch and is designated by $Y_F$. Again in Fig. 5.21, element 8 is connected to element 3 on the front of the armature, the connections being made at the commutator segment. Hence, $Y_F = 8 - 3 = 5$.

Alternatively, the front pitch may be defined as the distance (in terms of armature conductors) between the second conductor of one coil and the first conductor of the next coil which are connected together at the front i.e. commutator end of the armature. Both front and back pitches for lap and wave-winding are shown in Fig. 5.18 and 5.19.

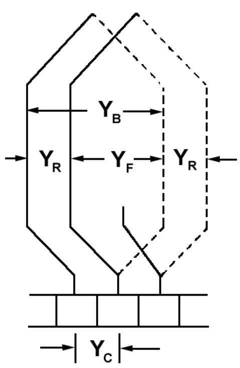



*Fig.5.18*

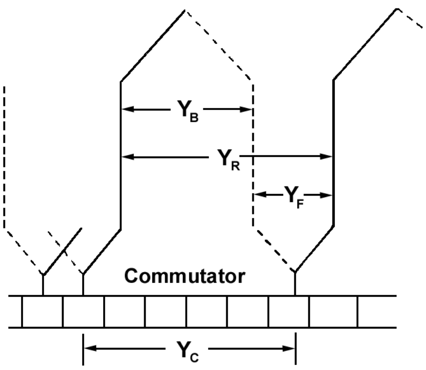



*Fig.5.19*

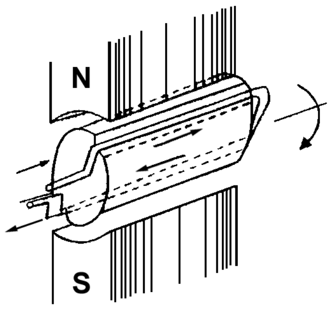



*Fig.5.20*

**Resultant Pitch ($Y_R$)**

It is the distance between the beginning of one coil and the beginning of the next coil to which it is connected (Fig. 5.18 and 5.19).

As a matter of precaution, it should be kept in mind that all these pitches, though normally stated in terms of armature conductors, are also sometimes given in terms of armature slots or commutator bars because commutator is, after all, an image of the winding.

**Commutator Pitch ($Y_C$)**

It is the distance (measured in commutator bars or segments) between the segments to which the two ends of a coil are connected. From Fig. 5.18 and 5.19 it is clear that for lap winding, $Y_C$ is the *difference* of $Y_B$ and $Y_F$ whereas for wave winding it is the *sum* of $Y_B$ and $Y_F$. Obviously, commutator pitch is equal to the number of bars between coil leads. In general, equals the 'plex' of the lap-wound armature. Hence, it is equal to 1, 2, 3, 4 etc. for simplex -, duplex, triplex- and quadruplex etc. lap-windings.

**Single-layer Winding**

It is that winding in which one conductor or one coil side is placed in each armature slot as shown in Fig. 5.20. Such a winding is not much used.

**Two-layer Winding**

In this type of winding, there are two conductors or coil sides per slot arranged in two layers. Usually, one side of every coil lies in the upper half of one slot and other side lies in the lower half of some other slot at a distance of approximately one pitch away (Fig. 5.21). The transfer of the coil from one slot to another is usually made in a radial plane by means of a peculiar bend or twist at the back end as shown in Fig. 5.22. Such windings in which two coil sides occupy each slot are most commonly used for all medium-sized machines. Sometimes 4 or 6 or 8 coil sides are used in each slot in several layers because it is not practicable to have too many slots (Fig. 5.23). The coil sides lying at the upper half of the slots are numbered odd i.e. 1, 3, 5, 7 etc. while those at the lower half are numbered even i.e. 2, 4, 6, 8 etc.

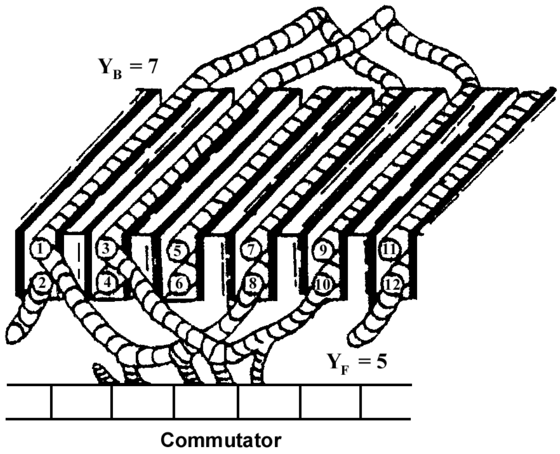



*Fig.5.21*

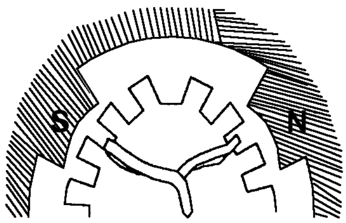



*Fig.5.22*

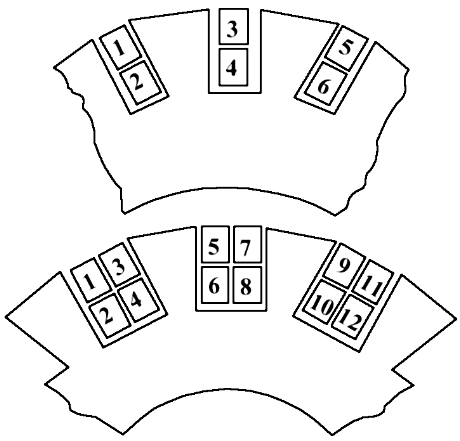



*Fig.5.23*

**Lap and Wave Windings**

Two types of windings mostly employed for drum-type armatures are known as Lap Winding and Wave Winding. The difference between the two is merely due to the different arrangement of the end connections at the front or commutator end of armature. Each winding can be arranged progressively or retrogressively and connected in simplex, duplex and triplex. The following rules, however, apply to both types of the windings :

i. The front pitch and back pitch are each approximately equal to the pole-pitch i.e. windings should be full-pitched. This results in increased e.m.f. round the coils. For special purposes, fractional-pitched windings are deliberately used.
ii. Both pitches should be odd, otherwise it would be difficult to place the coils (which are former-wound) properly on the armature. For example, if $Y_B$ and $Y_F$ were both even, the all the coil sides and conductors would lie either in the upper half of the slots or in the lower half. Hence, it would become impossible for one side of the coil to lie in the upper half. Hence, it would become impossible for one side of the coil to lie in the upper half of one slot and the other side of the same coil to lie in the lower half of some other slot.
iii. The number of commutator segments is equal to the number of slots or coils (or half the number of conductors) because the front ends of conductors are joined to the segments in pairs.
iv. The winding must close upon itself i.e. if we start from a given point and move from one coil to another, then all conductors should be traversed and we should reach the same point again without a break or discontinuity in between.

**Uses of Lap and Wave Windings**

The advantage of the wave winding is that, for a given number of poles and armature conductors, it gives more e.m.f. than the lap winding. Conversely, for the same e.m.f., lap winding would require large number of conductors which will result in higher winding cost and less efficient utilization of space in the armature slots. Hence, wave winding is suitable for small generators especially those meant for 500-600 V circuits.

Another advantage is that in wave winding, equalizing connections are not necessary whereas in a lap winding they definitely are. It is so because each of the two paths contains conductors lying under all the poles whereas in lap-wound armatures, each of the parallel paths contains conductors which lie under one pair of poles. Any inequality of pole fluxes affects two paths equally, hence their induced e.m.fs. are equal. In lap-wound armatures, unequal voltages are produced which set up a circulating current that produces sparking at brushes.

However, when large currents are required, it is necessary to use lap winding, because it gives more parallel paths.

Hence, lap winding is suitable for comparatively low-voltage but high-current generators whereas wave-winding is used for high-voltage, low-current machines.

**Simplex Wave Winding**

From Fig. 5.24, it is clear that in lap winding, a conductor (or coil side) under one pole is connected at the back to a conductor which occupies an almost corresponding position under the next pole of *opposite* polarity (as conductors 3 and 12). Conductor No.12 is then connected to conductor No. 5 under the *original* pole but which is a little removed

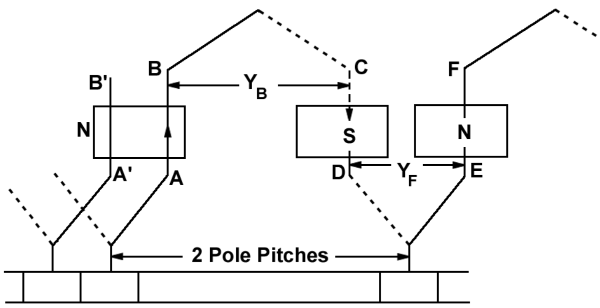



*Fig.5.24*

from the initial conductor No.3. If, instead of returning to the same *N*-pole, the conductor No.12 were taken *forward* to the next *N*-pole, it would make no difference so far as the direction and magnitude of the e.m.f. induced in the circuit are concerned.

As shown in Fig.5.24, conductor AB is connected to CD lying under S-pole and then to EF under the next N-pole. In this way, the winding progresses, passing successively under every N-pole and S-pole till it returns to a conductor A' B' lying under the original pole. Because the winding progresses in one direction round the armature in a series of 'waves', it is known as wave winding.

If, after passing once round the armature, the winding falls in a slot to the left of its starting points (as A' B' in Fig.5.24) then the winding is said to be retrogressive. If, however, it falls one slot to the right, then it is progressive.

Assuming a 2-layer winding and supposing that conductor AB lies in the upper half of the slot, then going once round the armature, the winding ends at A' B' which must be at the upper half of the slot at the left or right. Counting in terms of *conductor*, it means that AB and A' B' differ by two conductors (although they differ by one slot).

From the above, we can deduce the following relations. If P = No. of poles, then

$$\left.\begin{array}{l} Y_B = \text{back pitch} \\ Y_F = \text{front pitch} \end{array}\right\} \text{nearly equal to pole pitch}$$

then $$Y_A = \frac{Y_B + Y_F}{2} = \text{average pitch ; } Z = \text{total No. of conductors or coil sides}$$

Then, $$Y_A \times P = Z \pm 2 \qquad Y_A = \frac{Z \pm 2}{P}$$

Since P is always even and $Z = PY_A \pm 2$, hence Z must always be even. Put in another way, it means that $\frac{Z \pm 2}{P}$ must be an even integer.

The plus sign will give a progressive winding and the negative sign a retrogressive winding.

**Points to Note :**

1. Both pitches $Y_B$ and $Y_F$ are odd and of the same sign.
2. Back and front pitches are nearly equal to the pole pitch and may be equal or differ by 2, in which case, they are respectively one more or one less than the average pitch.

3. Resultant pitch $Y_R = Y_F + Y_B$.
4. Commutator pitch, $Y_C = Y_A$ (in lap winding $Y_C = \pm 1$).

Also, $$Y_C = \frac{\text{No. of Commutator bars } \pm 1}{\text{No. of pair of poles}}$$

5. The average pitch which must be an integer is given by

$$Y_A = \frac{Z \pm 2}{P} = \frac{\frac{Z}{2} + 1}{P/2} = \frac{\text{No. of Commutator bars } \pm 1}{\text{No. of pair of poles}}$$

It is clear that for $Y_A$ to be an integer, there is a restriction on the value of Z. With Z = 32, this winding is impossible for a 4-pole machine (though lap winding is possible). Values of Z = 30 or 34 would be perfectly all right.

6. The number of coils i.e. $N_C$ can be found from the relation.

$$N_C = \frac{PY_A \pm 2}{2}$$

This relation has been found by rearranging the relation gives in (5) above.

7. It is obvious from (5) that for a wave winding, the number of armature conductors with 2 either added or subtracted must be a multiple of the number of poles of the generator. This restriction eliminates many even numbers which are unsuitable for this winding.
8. The number of armature parallel paths = 2m where m is the multiplicity of the winding.

**Simplex Lap-winding**

It is shown in Fig. 5.25 which employs single-turn coils. In lap winding, the finishing end of one coil is connected to a commutator segment and to the starting end of the adjacent coil situated under the same pole and so on, till all the coils have been connected. This type of winding derives its name from the fact it doubles or laps back with its succeeding coils.

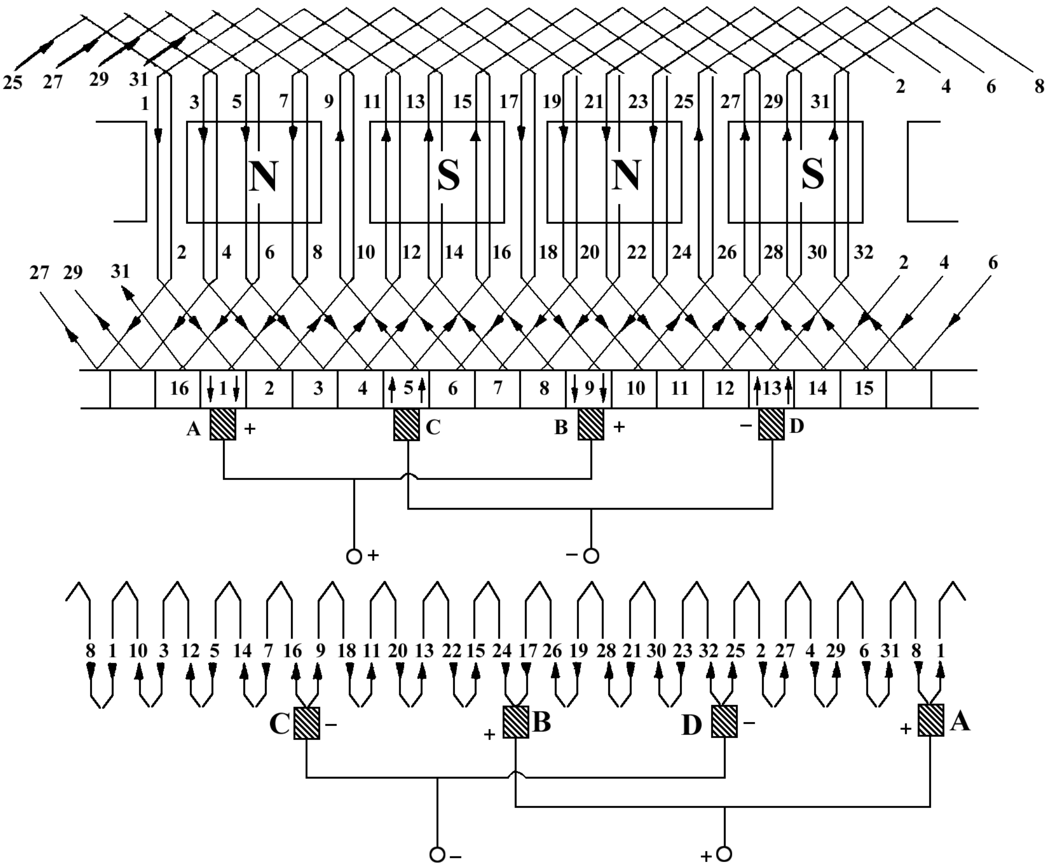


*Fig.5.25*

Following points regarding simplex lap winding should be carefully noted :

1. The back and front pitches are odd and of opposite sign. But they cannot be equal. They differ by 2 or some multiple thereof.

2. Both $Y_B$ and $Y_F$ should be nearly equal to a pole pitch.
3. The average pitch $Y_A = \dfrac{Y_B + Y_F}{2}$. It equals pole pitch $= \dfrac{Z}{P}$.
4. Commutator pitch $Y_C = \pm 1$. (In general, $Y_C = \pm m$)
5. Resultant pitch $Y_R$ is even, being the arithmetical difference of two odd numbers, i.e., $Y_R = Y_B - Y_F$
6. The number of slots for a 2-layer winding is equal to the number of coils (i.e. half the number of coil sides). The number of commutator segments is also the same.
7. The number of parallel paths in the armature = *mP* where *m* is the multiplicity of the winding and P the number of poles.

Taking the first condition, we have $Y_B = Y_F \pm 2$.

a. If $Y_B > Y_F$ i.e. $Y_B = Y_F + 2$, then we get a progressive or right-handed winding i.e. a winding which progresses in the clockwise direction as seen from the commutator end. In this case, obviously, $Y_C = +1$.
b. If $Y_B < Y_F$ i.e. $Y_B = Y_F - 2$, then we get a retrogressive or left-handed winding i.e. one which advances in the anti-clockwise direction when seen from the commutator side. In this case, $Y_C = -1$.
c. Hence, it is obvious that

$$\left.\begin{aligned} Y_F &= \frac{Z}{P} - 1 \\ Y_B &= \frac{Z}{P} + 1 \end{aligned}\right] \text{ for progressive winding and } \left.\begin{aligned} Y_F &= \frac{Z}{P} + 1 \\ Y_B &= \frac{Z}{P} - 1 \end{aligned}\right] \text{ for retrogressive winding}$$

Obviously, Z/P must be even to make the winding possible.

**Generated E.M.F. or E.M.F. Equation of a Generator**

Let Z = total number of armature conductors
= No. of slots × No. of conductors/slot
P = No. of generator poles
A = No. of parallel paths in armature
N = armature rotation in revolutions per minute (r.p.m.)
E = e.m.f. induced in any parallel path in armature

Generated e.m.f. $E_g$ = e.m.f. generated in any one of the parallel paths i.e. E.

$$\text{Average e.m.f. generated/conductor} = \frac{d\Phi}{dt} \text{ volt} \quad (\because n = 1)$$

Now, flux cut/conductor in one revolution $d\Phi = \Phi P$ Wb

No. of revolutions/second = N/60 $\therefore$ time for one revolution, dt = 60/N second

Hence, according to Faraday's Laws of Electromagnetic Induction,

$$\text{E.M.F. generated/conductor} = \frac{d\Phi}{dt} = \frac{\Phi PN}{60} \text{ volt}$$

**For a simplex wave-wound generator**

No. of parallel paths = 2

No. of conductors (in series) in one path = Z/2

$$\therefore \quad \text{E.M.F. generated/path} = \frac{\Phi PN}{60} \times \frac{Z}{2} = \frac{\Phi ZPN}{120} \text{ volt}$$

**For a simplex lap-wound generator**

No. of parallel paths = P

No. of conductors (in series) in one path = Z/P

$$\therefore \quad \text{E.M.F. \# r [generated/path 6} = \frac{\Phi PN}{60} \times \frac{Z}{P} = \frac{\Phi ZN}{60} \text{ volt}$$

$$\text{In general generated e.m.f. } E_g = \frac{\Phi ZN}{60} \times \left(\frac{P}{A}\right) \text{ volt}$$

Where A = 2 for simplex wave-winding
= P-for simplex lap-winding

Also, $$E_g = \frac{1}{2\pi} \cdot \left(\frac{2\pi N}{60}\right) \Phi Z \left(\frac{P}{A}\right) = \frac{\omega \Phi Z}{2\pi}\left(\frac{P}{A}\right) \text{volt - } \omega \text{ in rad/s}$$

For a given d.c. machine, Z, P and A are constant. Hence, putting $K_a = ZP/A$, we get

$$E_g = K_a \Phi N \text{ volts -- where N is in r.p.s.}$$

**TYPES OF GENERATORS**

Generators are usually classified according to the way in which their fields are excited. Generators may be divided into

a) separately-excited generators and
b) self-excited generators.

**a. Separately-excited**

Generators are those whose field magnets are energized from an independent external source of d.c. current. It is shown diagrammatically in Fig. 5.26.

**b. Self-excited**

Generators are those whose field magnets are energized by the current produced by the generators themselves. Due to residual magnetism, there is always present some flux in the poles. When the armature is rotated, some e.m.f. and hence some induced current is produced which is partly or fully passed through the field coils thereby strengthening the residual pole flux.

There are three types of self-excited generators named according to the manner in which their field coils (or windings) are connected to the armature.

**i. Shunt wound**

The field windings are connected across or in parallel with the armature conductors and have the full voltage of the generator applied across them (Fig. 5.27).

**ii. Series Wound**

In this case, the field windings are joined in series with the armature conductors. As they carry full load current, they consist of relatively few turns of thick wire or strips. Such generators are rarely used except for special purposes such as boosters etc.

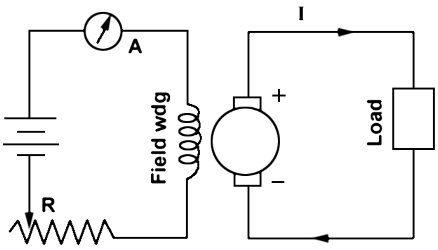



*Fig.5.26, Separately excited Generator*

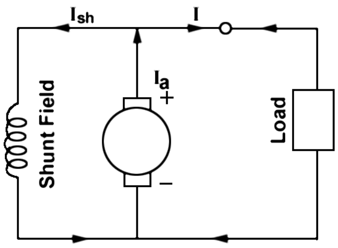



*Fig.5.27, Shunt wound Generator*

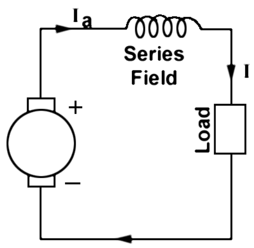



*Fig.5.28, Series wound generator.*

**iii. Compound Wound**

It is a combination of a few series and a few shunt windings and can be either short-shunt or long-shunt as shown in Fig. 26.44 (a) and (b) respectively. In a compound generator, the shunt field is stronger than the series field. When series field aids the shunt field, generator is said to be commutatively-compounded. On the other hand if series field opposes the shunt field, the generator is said to be differentially compounded. Various types of d.c. generators have been shown separately in Fig. 26.45.

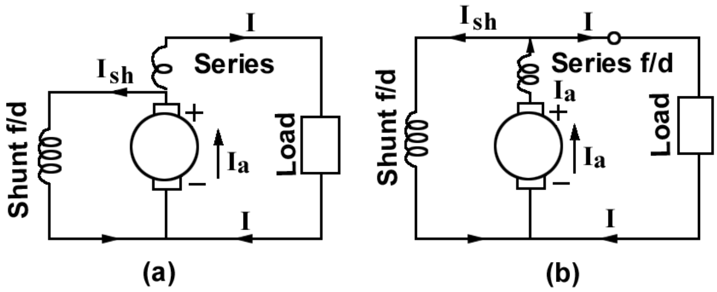



*Fig.5.29, Compound wound generator.*

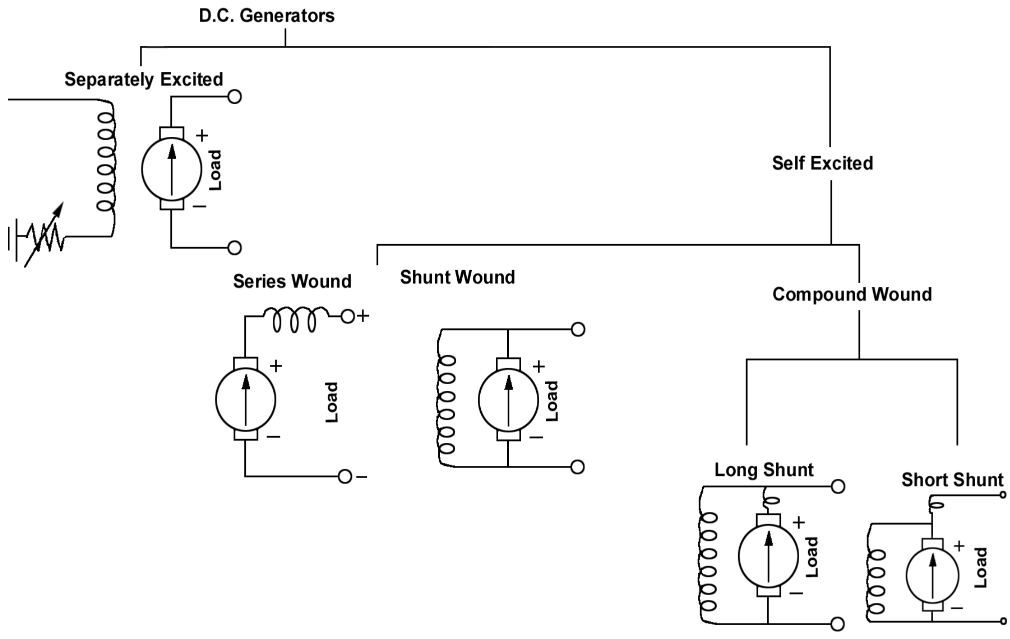



*Fig.5.30*

## ARMATURE REACTION IN DC GENERATORS

Armature reaction is the effect of magnetic flux set up by armature current upon the distribution of flux under the main poles.

Fig. 5.31 shows a 2-pole d.c. generator. When there is *no load* connected to the generator, the current in the armature conductors is zero. Under these conditions there exists in it only the m.m.f. of the main poles which produce the *main flux* $\phi$. This flux is distributed symmetrically with respect to the polar axis, that is, the centre line of the north and the south poles. The direction of $\Phi_M$ is shown by an arrow. The **magnetic neutral axis** or plane (MNA) *is a plane perpendicular to the axis of the flux.*

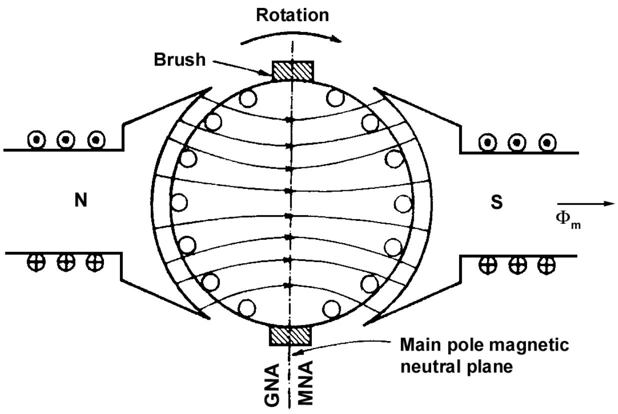



*Fig.5.31. Main pole magnetic flux distribution.*

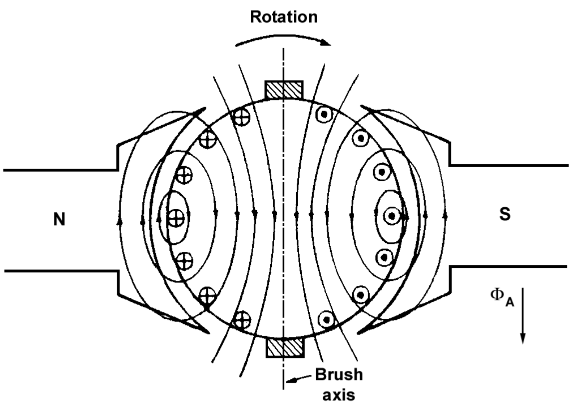



*Fig.5.32. Armature flux distribution.*

The MNA coincides with the geometrical neutral axis or plane (GNA). *Brushes are always placed along MNA. Hence MNA is also called the* ***axis of commutation.***

Fig. 5.32 shows armature conductors carrying current with no current in field coils. The direction of current in the armature conductors may be determined by Fleming's right-hand rule. The current flows in the same direction in all the conductors lying under one pole. The direction of flux produced by armature conductors may be determined by cork-screw rule. The conductors on the left-hand side of the armature carry current in the direction into the paper. The flux produced by current in these armature conductors in shown in Fig. 5.32.

These conductors combine their m.m.fs to produce a flux through the armature in the downward direction. Similarly, the conductors on the right-hand side of the armature carry current in the direction out of the paper. These conductors also combine their m.m.fs to produce a flux through the armature in the downward direction. Thus, the conductors on both sides of the armature combine their m.m.fs in such a manner as to send a flux through the armature in the downward direction. This flux $\Phi_A$ is represented by an arrow as shown in Fig. 6.12. The armature flux produced is analogous to that produced in the equivalent iron-cored solenoid with its axis along the brush axis.

Fig. 5.33 shows the condition when the field current and armature current are acting simultaneously. This occurs when the generator is on load. Now there are two fluxes inside the machine, one produced by the main field poles of the generator and the other by the current in the armature conductors. These two fluxes now combine to form a resultant flux $\Phi_R$ as shown in Fig. 5.33.

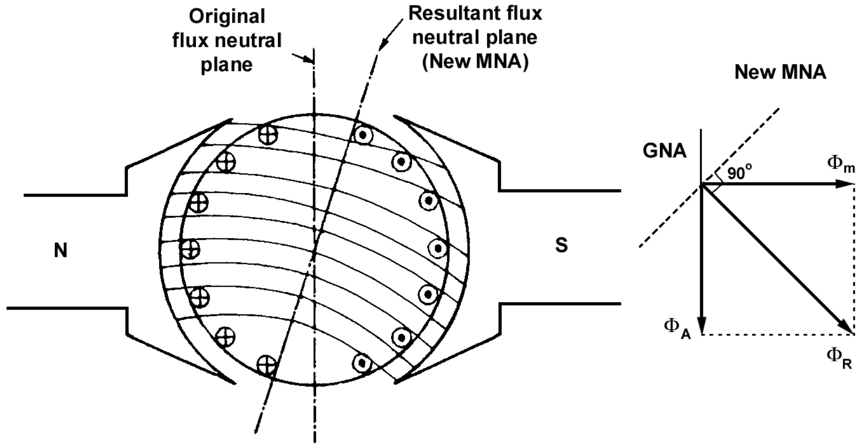



*Fig.5.33. Resultant flux distribution.*

It is seen that the field flux entering the armature is not only *shifted* but also distorted. The distortion produces crowding of the flux (increase in the flux density) in the upper pole tip in the N-pole and in the lower pole tip in the S-pole. Similarly, there is a reduction of flux (decreased flux density) in the lower tip of the N-pole and in the upper pole tip of the S-pole. The direction of the resultant flux has shifted in the direction of rotation of the generator. Since the MNA is always perpendicular to the axis of the resultant flux, the MNA is also shifted.

Because of the nonlinear behaviour and saturation of the core, the increase in flux in one pole tip is less than the decrease in flux in the other pole tip consequently, the main pole flux is decreased. Since $E_g = kN\Phi$, the reduction in the field flux $\Phi$ decreases the terminal voltage of a generator with increased load.

**Effects of Armature Reaction**

The effects of armature reaction are summarized below :

1. Magnetic flux density is increased over one half of the pole and decreased over the other half. But the total flux produced by each pole is slightly reduced and, therefore, the terminal voltage is slightly reduced. The effect of total flux reduction by armature reaction is known as **demagnetizing effect.**
2. The flux wave is distorted and there is shift in the position of the magnetic neutral axis (MNA) in the direction of rotation for the generator and against the direction of rotation for the motor.
3. Armature reaction establishes a flux in the neutral zone (or commutating zone). Armature reaction flux in the neutral zone will induce conductor voltage that aggravates the commutation problem.

**COMMUTATION**

The currents induced in the armature conductors of a d.c. generator are alternating in nature. The commutation process involves the change from a generated alternating current to an externally applied direct current. These induced currents flow in one direction when the armature conductors are under north pole. They are in opposite direction when they are under south pole. As conductors pass out of the influence of north pole and enter the south pole, the current in them is reversed. The reversal of current takes place along the MNA or brush axis. Whenever a brush spans two commutator segments, the winding element connected to those segments is short circuited. By commutation we mean the change that take place in a winding element during the period of short circuit by a brush. These changes are shown in Fig. 5.34. For simplicity, consider a simple ring winding.

In position shown in Fig. 5.34 (a), the current I flowing towards the brush from the left-hand side passes round the coil in a clockwise direction.

In position shown in Fig. 5.34(b), this coil carries the same current in the same direction, but the coil is to short circuited by the brush.

In position shown in Fig. 5.34 (c), the brush makes contact with bars a and b, thereby short circuiting coil 1. The current is still I from the left-hand side and I from the right-hand side. It is seen that these two currents can reach the brush without passing through coil 1.

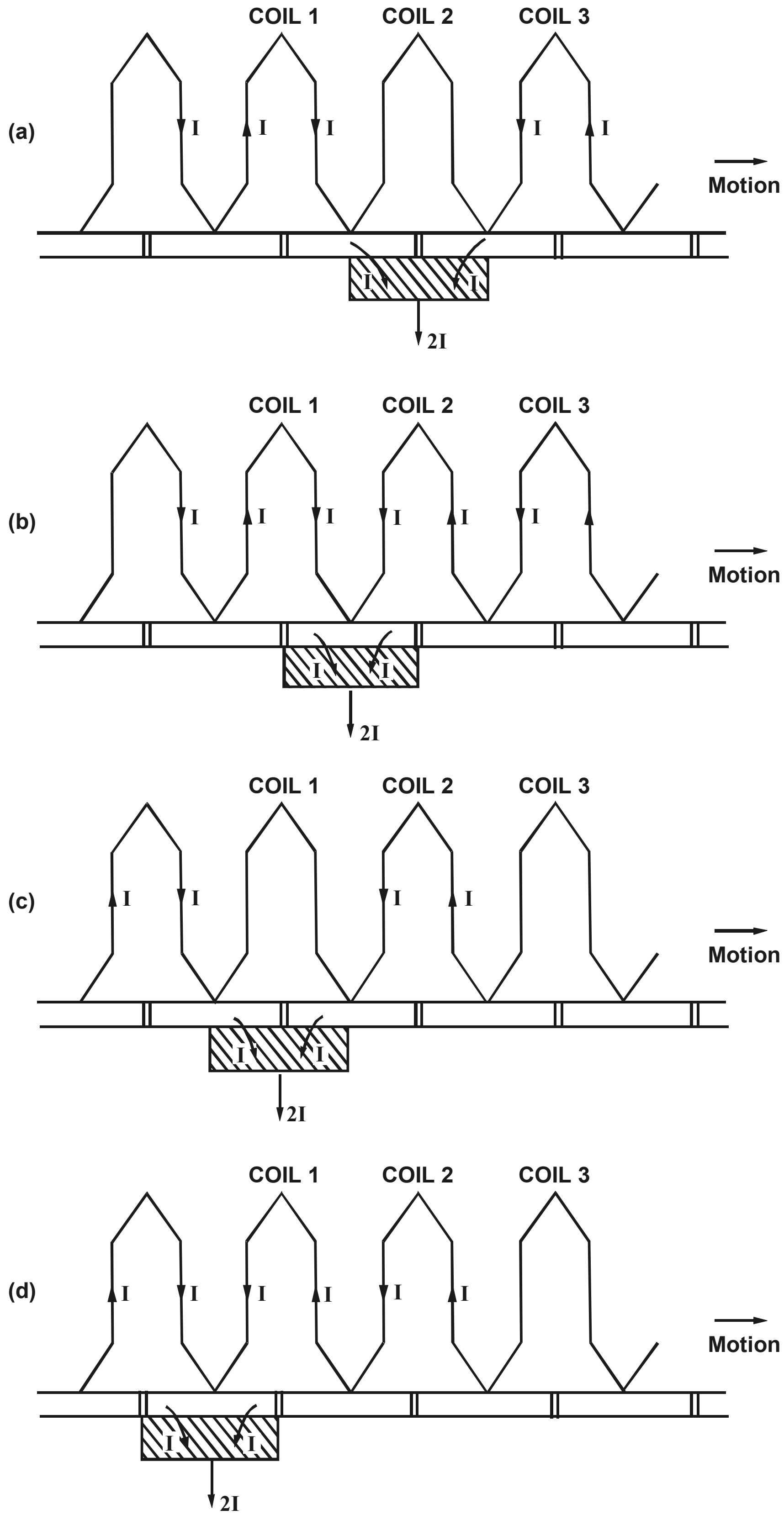



*Fig.5.34. Current collection at the commutator.*

Fig. 5.34(d), shows that bar b has just left the brush and the short circuit of coil 1 has ended. It is now necessary for the current I reacting the brush from the right-hand side in the anticlockwise direction.

From the above discussion it is seen that during the period of short circuit of an armature coil by a brush the current in that coil must be reversed and also brought up to its full value in the reversed direction. The time of short circuit is called the **period of commutation.**

Fig. 5.35 shows how the current in the short-circuited coil varies during the brief interval of the short circuit. Curve B shows that the current changes from + I to - I linearly in the commutation period. Such a commutation is called **ideal commutation** or **straight-line commutation.**

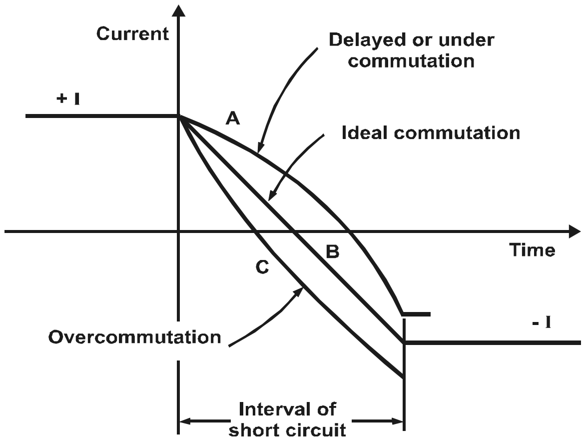



***Fig.5.35. Coil current as a function of time during commutation.***

If the current through coil 1 has not reached its full value in the position shown in Fig. 5.34 (d), then since coil 2 is carrying full current, the difference between the currents, through elements 2 and 1 has to jump from commutator bar b to the brush in the form of a spark. Thus, the **cause of sparking at the commutator** *is the failure of the current in the short-circuited elements to reach the full value in the reversed direction by the end of short circuit.*

This is known as under commutation or delayed commutation. The curve of current against time in such a case is shown in Fig. 5.35 by curve A. In ideal commutation (curve B) the current of the commutating coils changes linearly from + I to - I in the commutation period. Curve C represents overcommutation or accelerated commutation when the current reaches its final value with a zero rate of change at the end of the commutation period. Usually this will result in a satisfactory commutation.

In actual practice, the current in the short-circuited coil after commutation period does not reach its full value. This is due to the fact that the short-circuited coil offers self-inductance in addition to resistance. The rate of change of current is so great that the self-inductance of the coil sets up a back e.m.f. which opposes the reversal. Since the current in the coil has to change from + I to - I, the total change is 2I. If $t_c$ is the time of short circuit and L is the inductance of the coil (= self-inductance of the short-circuited coil + mutual inductances of the neighbouring coils), then the average value of the self induced voltage is

$$L\frac{di}{dt} = \frac{L \times 2I}{t_c} = \frac{2LI}{t_c}$$

This is called the **reactance voltage.**

The large voltage appearing between commutator segments to which the coil is connected causes sparking at the brushes of the machine the sparking of commutator is much harmful and it will damage both commutator surface and brushes. Its effect is cumulative which may lead to a short circuit of the machine with an arc round the commutator from brush to brush.

## METHODS OF IMPROVING COMMUTATION

There are three methods of obtaining sparkless commutation :

1. Resistance commutation.
2. Voltage commutation.
3. Compensating windings.

### Resistance Commutation

This method of improving commutation consists of using carbon brushes. This makes the contact resistance between commutator segments and brushes high. This high contact resistance has the tendency to force the current in the short-circuited coil to change according to the commutation requirements, namely, to reverse and then build up in the reversed direction.

**Voltage commutation**

In this method, arrangement is made to induce a voltage in the coil under-going commutation, which will neutralize the reactance voltage. This injected voltage is in opposition to the reactance voltage. If the value of the injected voltage is made equal to the reactance voltage, quick reversal of current in the short-circuited coil will take place and there will be sparkless commutation.

Two methods may be used to produce the injected voltage in opposition to the reactance voltage :
1. Brush shift.
2. Commutating poles or interpoles.

**Brush Shift**

The effect of armature reaction is to shift the magnetic neutral axis MNA in the direction of rotation for the generator and against the direction of rotation for the motor. Armature reaction establishes a flux on the neutral zone. A small voltage is induced in the commutating coil since it is cutting the flux.

**Commutating Poles or Interpoles**

Interpoles are narrow poles attached to the stator yoke, and placed exactly midway between the main poles. Interpoles are also called **commutating poles** or **compoles**. The interpole windings are connected in series with the armature, because the interpoles must produce fluxes that are directly proportional to the armature current. The armature and interpole m.m.fs are affected simultaneously by the same armature current. Consequently, the armature flux in the commutating zone which tends to shift the magnetic neutral axis, is neutralized by an appropriate component of interpole flux. The neutral plane is, therefore, fixed in position regardless of the load.

The interpoles must induce a voltage in the conductors undergoing commutation that is opposite to the voltage caused by the neutral-plane shift and reactance voltage. For a generator, the neutral plane shifts in the direction of rotation. Thus, the conductors undergoing commutation have the same polarity of the voltage as the pole they just left. To oppose this voltage, the interpoles must have the opposite flux, which is the flux of the main pole ahead according to the direction of rotation.

For a motor, the neutral plane shifts opposite to the direction of rotation, and the conductors undergoing commutation have the same flux as the main pole then are approaching. For opposing this voltage, the interpoles must have the same polarity as the previous main pole. Thus we have the following rules for the polarity of the interpoles:
1. For a *generator*, the polarity of the interpole must be the same as that of the next main pole further ahead in the direction of rotation.
2. For a *motor,* the polarity of an interpole is opposite to that of the next main pole in the direction of rotation. The polarity of interpoles is shown in Fig. 5.36.

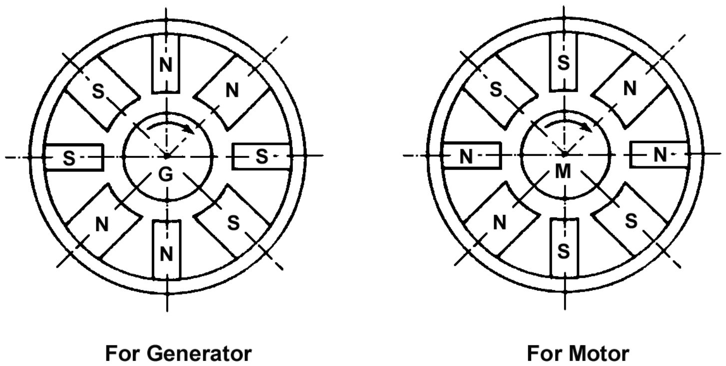



*Fig.5.36. Polarity of interpoles.*

It is to be noted that the interpoles serve only to provide sufficient flux to assure good commutation. They do not overcome the distortion of the flux resulting from cross-magnetizing m.m.f. of the armature.

The use of interpoles is very common to nearly all dc machines of more than 1 hp.

During severe overloads or rapidly changing loads the voltage between adjacent commutator segments may become very high. This may ionize the air around the commutator to the extent that it becomes sufficiently conductive. An arc is established from brush to brush. This phenomenon is known as flashover. This arc is sufficiently hot to melt the commutator segments. It should be extinguished quickly. In order to prevent flashover compensating windings are used.

**Compensating windings**

Compensating windings are the most effective means for eliminating the problems of armature reaction and flashover by balancing the armature m.m.f.. Compensating windings are placed in slots provided in pole faces parallel to the rotor (armature) conductors. These windings are connected in series with the armature windings. The direction of currents in the compensating winding must be opposite to that in the armature winding just below the pole faces. Thus, compensating winding produces an m.m.f. that is equal and opposite to the armature m.m.f.. In effect the compensating winding demagnetizes or neutralizes the armature flux produced by the armature conductors lying just under the pole faces. The flux per pole is then undisturbed by the armature flux regardless of the load conditions.

The major drawback with the compensating windings is that they are very costly. Their use can only be justified in the following special cases :

1. In large machines subject to heavy overloads or plugging.
2. In small motors subject to sudden reversal and high acceleration.

**CHARACTERISTICS OF DC GENERATORS**

At present time bulk of electrical energy is generated in the form of alternating current. DC generators are no more used in modern power systems. For the sake of continuity, the characteristics of dc generators are briefly given here. Characteristic is the graph between two dependent quantities.

Following are the three important characteristics of a dc generator :

1. **Magnetization Characteristic.** Magnetization characteristic gives the variation of generated voltage (or no-load voltage) with field current at a constant speed. It is also called no-load or open-circuit characteristic (O.C.C.).
2. **Internal Characteristic.** It is the plot between the generated voltage and load current.
3. **External Characteristic** or Load Characteristic
   It is a graph between the terminal voltage and load current at a constant speed.

**SEPARATELY EXCITED DC GENERATOR**

In the separately excited dc generator, the field winding is connected to a separate source of dc power. This source may be another dc generator, a battery, a diode rectifier, or a controlled rectifier.

The circuit for a separately excited dc generator on load is shown in Fig. 5.37.

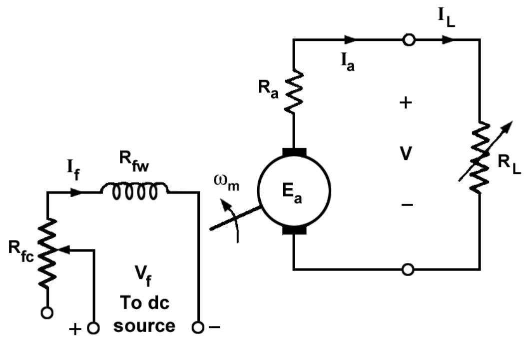



*Fig.5.37, Circuit model of a separately excited dc generator.*

Let the generator be driven at a constant speed by a prime mover. The field excitation ($I_f$) is adjusted to give rated voltage at no load. This value of voltage is kept constant throughout the operation considered.

Let $R_{fw}$ = resistance of the field winding
$R_{fc}$ = resistance of the field rheostat to control field current
$R_f$ = total field circuit resistance $= R_{fw} + R_{fc}$
$R_a$ = total resistance of the armature circuit
(including the brush-contact resistance)
$R_L$ = load resistance
$I_L$ = load current
$E_a$ = internal generated voltage
$V$ = terminal voltage
$I_a$ = armature current

The defining equations for the separately excited dc generator are as follows :

$V_f = R_f I_f$
$E_a = V + I_a R_a$
$E_a = K_a \Phi \omega_m$
$V = I_L R_L$
$I_a = I_L$

If there were no armature reaction, the generated voltage $V_0$ would be constant as shown by a straight line in Fig.5.38. Because of the demagnetizing effect of armature reaction there is a voltage drop $\Delta V_{AR}$. The internal characteristic ($E_a \sim I_L$) is shown in Fig. 5.38.

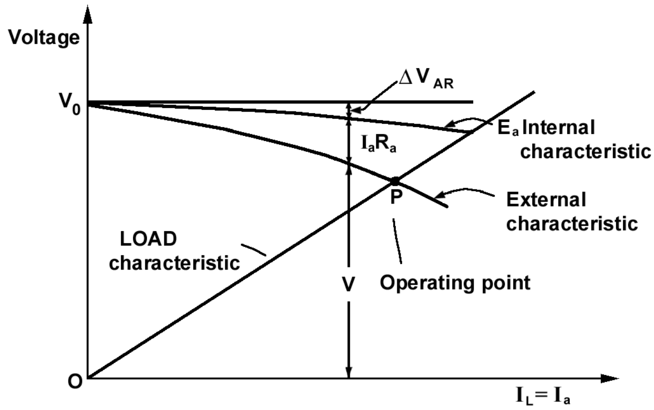



*Fig.5.38. Terminal characteristics of a separately excited dc generator.*

There is a voltage drop $I_a R_a$ across $R_a$. The generator external characteristic ($V \sim I_L$) defined by the relation

$$V = E_a - I_a R_a$$

is shown in Fig. 5.38. The point of intersection between the generator external characteristic and the load characteristic given by the relation $V = I_L R_L$ determines the operating point P. The operating point gives the operating values of terminal voltage V and terminal (load) current $I_L$.

**VOLTAGE BUILDUP IN SELF-EXCITED GENERATORS**

A self-excited dc generator supplies its own field excitation. A self-excited generator shown in Fig. 5.39 is known as a shunt generator because its field winding is connected in parallel with the armature. Thus, the armature voltage supplies the field current.

This generator will build up a desired terminal voltage. Assume that the generator in Fig. 5.39 has no load connected to it and the armature is driven at a certain speed by a prime mover. We shall study the conditions under which the voltage build up takes place. The voltage buildup in a dc generator depends upon the presence of a residual flux in the field poles of the generator. A small voltage $E_{ar}$ will be generated. It is given by

$$E_{ar} = K\Phi_{res}\,\omega \qquad (5.15.1)$$

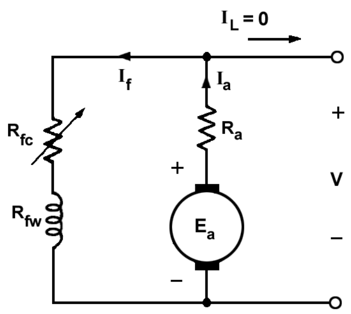



*Fig.5.39. Equivalent circuit of a shunt dc generator.*

This voltage is of the order of 1V or 2V. it causes a current $I_f$ to flow in the field winding of the generator. The field current is given by

$$I_f = \frac{V}{R_f} \qquad (5.15.2)$$

This field current produces a magnetomotive force in the field winding, which increases the flux. The increase in flux increases the generated voltage $E_a$. The increased armature voltage $E_a$ increases the terminal voltage V. With the increase in V, the field current $I_f = \frac{V}{R_f}$ increases further. This in turn increases $\Phi$ and consequently $E_a$ increases further. The process of voltage buildup continues. Fig. 5.40 shows the voltage buildup of a dc shunt generator.

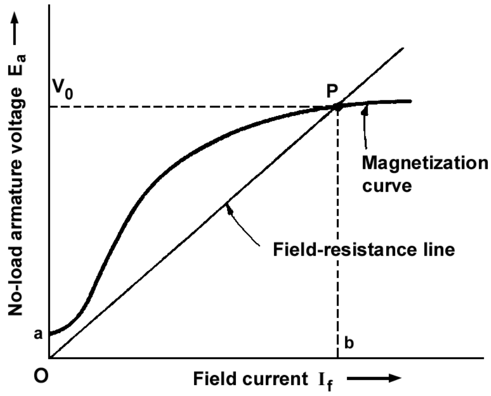



*Fig.5.40. Voltage build up of a dc shunt generator.*

The effect of magnetic saturation in the pole faces limits the terminal voltage of the generator to a steady-state value.

We have assumed that the generator is on no load during the buildup process. The following equations describe the steady-state operation :

$$I_a = I_f \qquad (5.15.3)$$

$$V = E_a - I_a R_a = E_a - I_f R_a$$

**Since the field current $I_f$ curve in a shunt generator is very small, the voltage drop $I_f R_a$ can be neglected,**

and $\quad V = E_a. \qquad (5.15.4)$

The $E_a$ versus $I_f$ curve is the magnetization curve shown in Fig. 5.40.

For the field circuit

$$V = I_f R_f \qquad (5.15.5)$$

The straight line given by

$$V = I_f R_f$$

is called the **field-resistance line.**

The field-resistance line is a plot of the voltage $I_f R_f$ across the field circuit versus the field current $I_f$. The slope of this line is equal to the resistance of the field circuit.

The solution of Equations (5.15.4) & (5.15.5) gives the no-load terminal voltage $V_0$ of the generator. Thus, the intersection point P of the magnetization curve and the field-resistance line gives the no-load terminal voltage $V_0$ (= bP) and the corresponding field current (Ob). Normally, in the shunt generator the voltage builds up to the value given by the point P. At this point $E_a = I_f R_f = V_0$.

If the field current corresponding to point P is increased further, there is no further increase in the terminal voltage.

Fig.5.41 shows the voltage buildup in the dc shunt generator for various field circuit resistances.

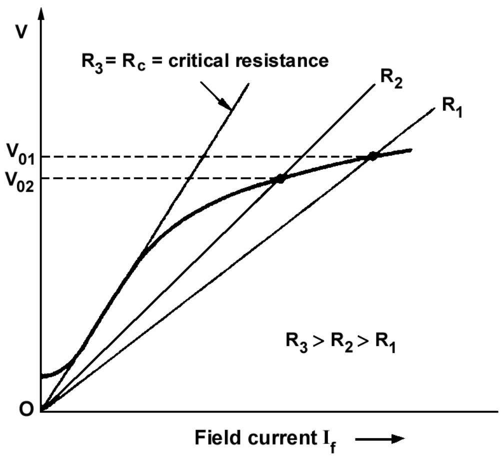



*Fig.5.41. Effect of field resistance on no-load voltage.*

A decrease in the resistance of the field circuit reduces the slope of the field-resistance line resulting in a higher voltage. If the speed remains constant. An increase in the resistance of the field circuit increases the slope of the field resistance line, resulting in a lower voltage. If the field circuit resistance is increased to $R_c$ which is termed as the critical resistance of the field, the field resistance line becomes a tangent to the initial part of the magnetization curve. When the field resistance is higher than this value, the generator fails to excite.

Fig. 5.42 shows the variation of no-load voltage with fixed $R_f$ and variable speed of the armature.

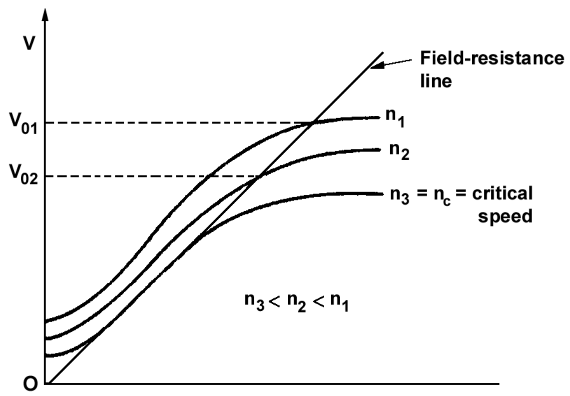



*Fig.5.42. Variation of no-load voltage with speed.*

The magnetization curve varies with the speed and its ordinate for any field current is proportional to the speed of the generator. If the field resistance is kept constant and the speed is reduced, all the points on the magnetization curve are lowered, and the point of intersection of the magnetization curve and the field resistance line moves downwards. At a particular speed, called the critical speed, the field-resistance line becomes tangential to the magnetization curve. Below the critical speed the voltage will not build up.

In brief, the following conditions must be satisfied for voltage build up in a self-excited dc generator.

1. There must be sufficient residual flux in the field poles.
2. The field terminals should be connected in such a way that the field current increases flux in the direction of residual flux.
3. The field circuit resistance should be less than the critical field circuit resistance.

If there is no residual flux in the field poles, disconnect the field from the armature circuit and apply a dc voltage to the field winding. This process is called **flashing the field.** It will induce some residual flux in the field poles.

**CHARACTERISTICS OF COMPOUND DC GENERATORS**

The voltage-current characteristics of compound generators are shown in Fig. 5.43.

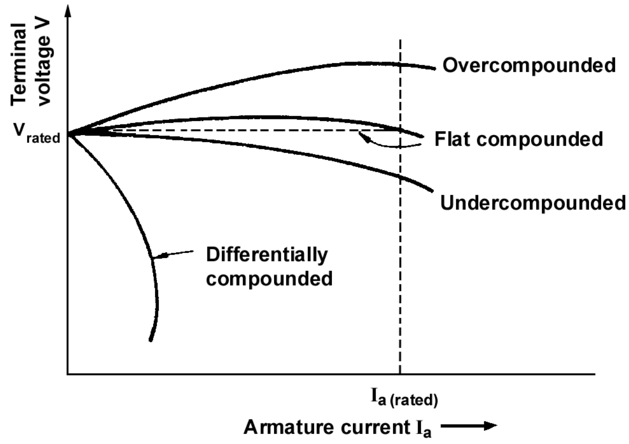



*Fig.5.43. Voltage-current characteristics of compound dc generators.*

Depending upon the number of series field turns, the cumulatively compounded generators may be overcompounded, flat compounded, and undercompounded. For an overcompounded generator full-load terminal voltage is higher than the no-load terminal voltage. For a flat compounded (or level compounded) generator the terminal voltage at full load is equal to the no-load terminal voltage. In an undercompounded generator the terminal voltage at full load is less than the no-load terminal voltage.

In differential compounded generators, the terminal voltage drops very quickly with increasing armature current.

■■■

# CHAPTER : 6
# ALTERNATINGCURRENTGENERATORS

Alternating current generators are generally of two types those designed for operation over a wide variable speed and variable frequency range (frequency wild generators), and those designed for constant speed and constant frequency operation (constant speed generators).

**FREQUENCYWILDSYSTEMS**

A frequency-wild system is one in which the frequency of its generator voltage output is permitted to vary with the rotational speed of the generator. Although such frequency variations are not suitable for the direct operation of all types of a.c. consumer equipment, the output can (after constant voltage regulation) be applied directly to resistive load circuits such as electrical de-icing systems, for the reason that resistance to alternating current remains substantially constant, and is independent of frequency.

**Generator Construction**

A typical frequency wild generator is utilized for the supply of heating current to a turbo propeller engine de-icing (Fig. 6.1) system. It has a three phase output of 22 KVA at 208 volts and it supplies full load at this voltage through a frequency range of 280 to 400 Hz. Below 280 Hz the field current is limited and the output relatively reduced. The generator consists of two major assemblies, a fixed stator assembly in which the current is induced and a rotating assembly referred to as the rotor. The stator assembly is made up of high permeability laminations and is damped in a main housing by an end frame having an integral flange for mounting the generator at the corresponding drive outlet of an engine driven gear box. The stator winding is star connected, the star or neutral point being made by linking three ends of the winding and connecting it to ground. The other three ends of the winding are brought out to a three way output terminal box mounted on the end frame of the generator. Three small current transformers are fitted into the terminal box and form part of a protection system known as a Merz-Price system.

The rotor assembly has six salient poles of laminated construction; their series connected field windings terminate at two slip rings secured at one end of the rotor shaft. Three spring loaded brushes are equi-spaced on each slip ring and are contained within a brush gears housing which also forms a being support for the rotor. The brushes are electrically connected to d.c. input terminals housed in an excitation terminal box mounted above the brush gear housing. The terminal box also houses capacitors which are connected between the terminals and frame to suppress interference, in the reception of radio signals. At the drive end, the rotor shaft in servated and an oil seal, housed in a carrier plate bolted to the main housing is fitted. Over the shaft to prevent the entry of oil from the diving source into the main housing.

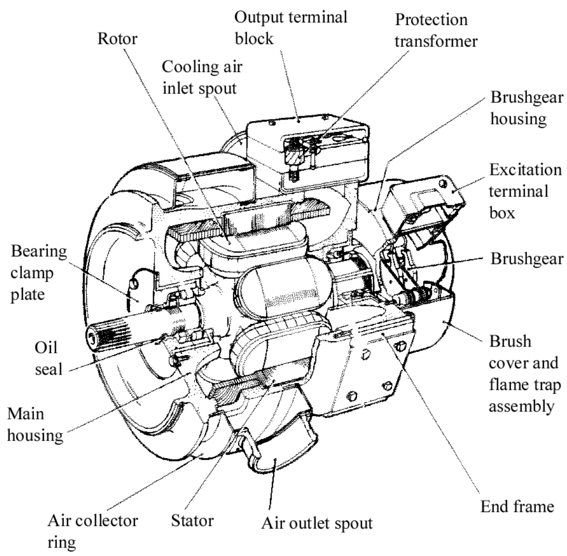



*Fig. 6.1, Frequency-wild generator*

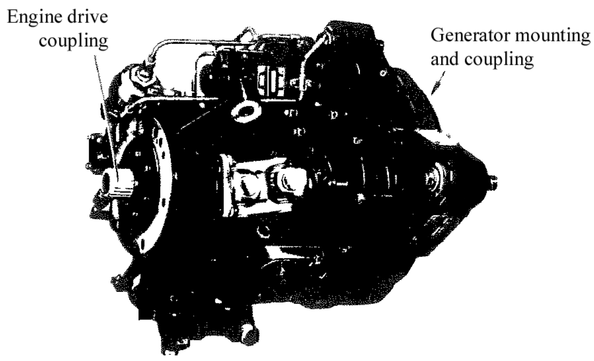



*Fig. 6.2, Constant speed drive unit.*

The generator is cooled by ram air passing into the main housing via an inlet spout at the slipring end, the air escaping from the main housing through ventilation slots at the drive ends.

## CONSTANT FREQUENCY SYSTEMS

In the development of electrical power supply systems notably for large aircraft, the idea was conceived of an all a.c. system i.e. a primary, generating system to meet all a.c. supply requirements, in particular those of numerous consumer services dependent on constant frequency, to allow for parallel generator operation, and to meet d.c. supply requirements via transformer and rectifier system.

A constant frequency is inherent in an a.c. system only if the generator is driven at a constant speed. The engines cannot be relied upon to do this directly and, as we have already learned, if a generator is connected directly to the accessory drive of an engine the output frequency will vary with engine speed. Some form of conversion equipment is therefore required and the type most widely adopted utilizes a transmission device interposed between the engine and generator, and which incorporates a variable-ratio drive mechanism. Such a mechanism is referred to as a constant-speed drive (CSD) unit.

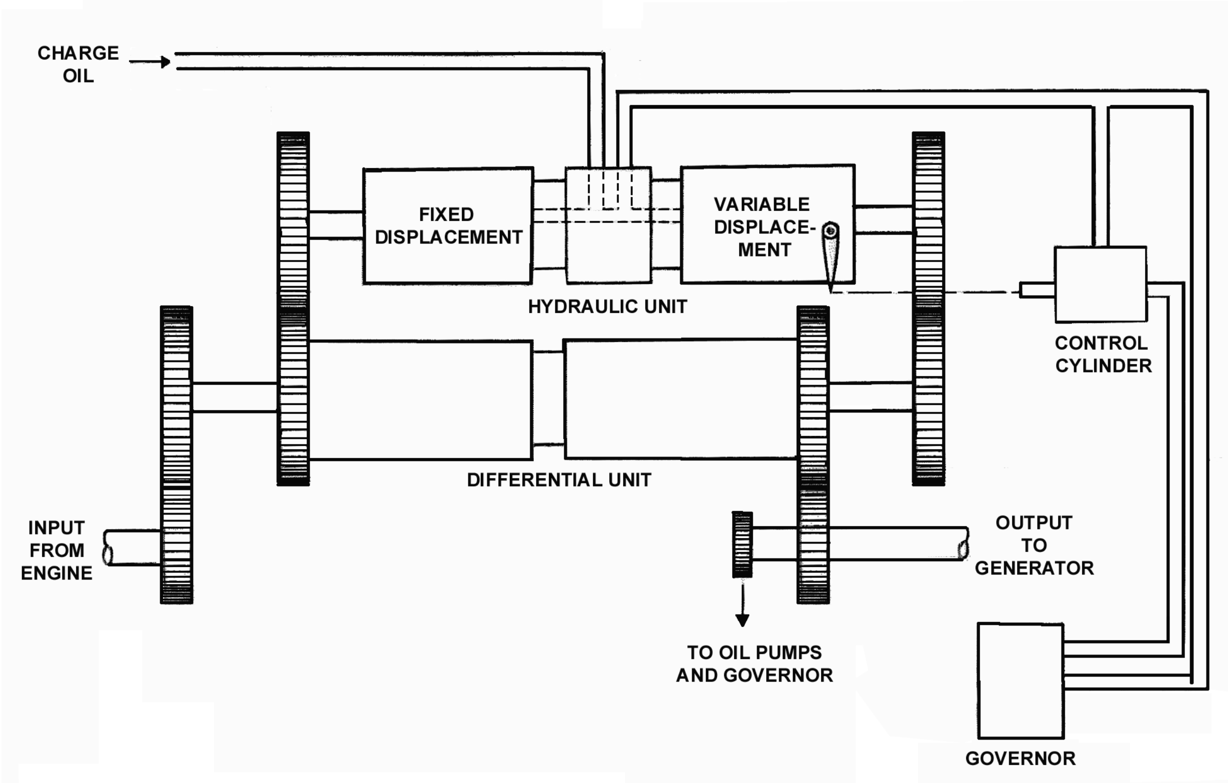



*Fig. 6.3, Basic arrangement of a CSD unit.*

The unit employs a hydromechanical variable-ratio drive which in its basic form consists of a variable displacement hydraulic unit, a fixed displacement hydraulic unit and a differential gear. The power used to drive the generator is controlled and transmitted through the combined effects of the three units. Oil for system operation is supplied from a reservoir via charge pumps within the unit, and a governor.

### Constant Frequency Generator Construction

A typical constant frequency generator consists of three principal components : a.c. exciter which generates the power for the generator field; rotating rectifier assembly mounted on and rotating with the rotor shaft to convert the exciter output to d.c. and the main generator. All three component are contained with in a cast Al casing made up of an end bell section and a stator frame section; both sections are secured externally by screws. A mounting flange, which is an integral part of the stator frame, carries twelve slots reinforced by steel inserts, and key hole shaped to facilitate attachment of the generator to the mounting studs of the constant speed drive unit.

The exciter which is located in the end bell section of the generator casing, comprises on stator and a three phase star wound rotor, or exciter armature. The exciter armature is mounted on the same shaft as the main generator rotor and the output from its three phase windings is fed to the rotating rectifier assembly.

The rotating rectifier assembly supplies excitation current to the main generator rotor field coils, and since together with the a.c. exciter they replace the conventional brushes and slip rings, they thereby eliminate the problems associated with them. The assembly is contained with in a tubular insulator located in the hollow shaft on which the exciter and main generator rotor are mounted; located in this manner they are close to the axis of rotation and are not therefore subjected to excessive centrifugal forces. A suppression capacitor is also connected in the rectifier circuit and is mounted at one end of the rotor shaft. Its purpose is to suppress voltage "spikes" created within the diodes under certain operating, conditions.

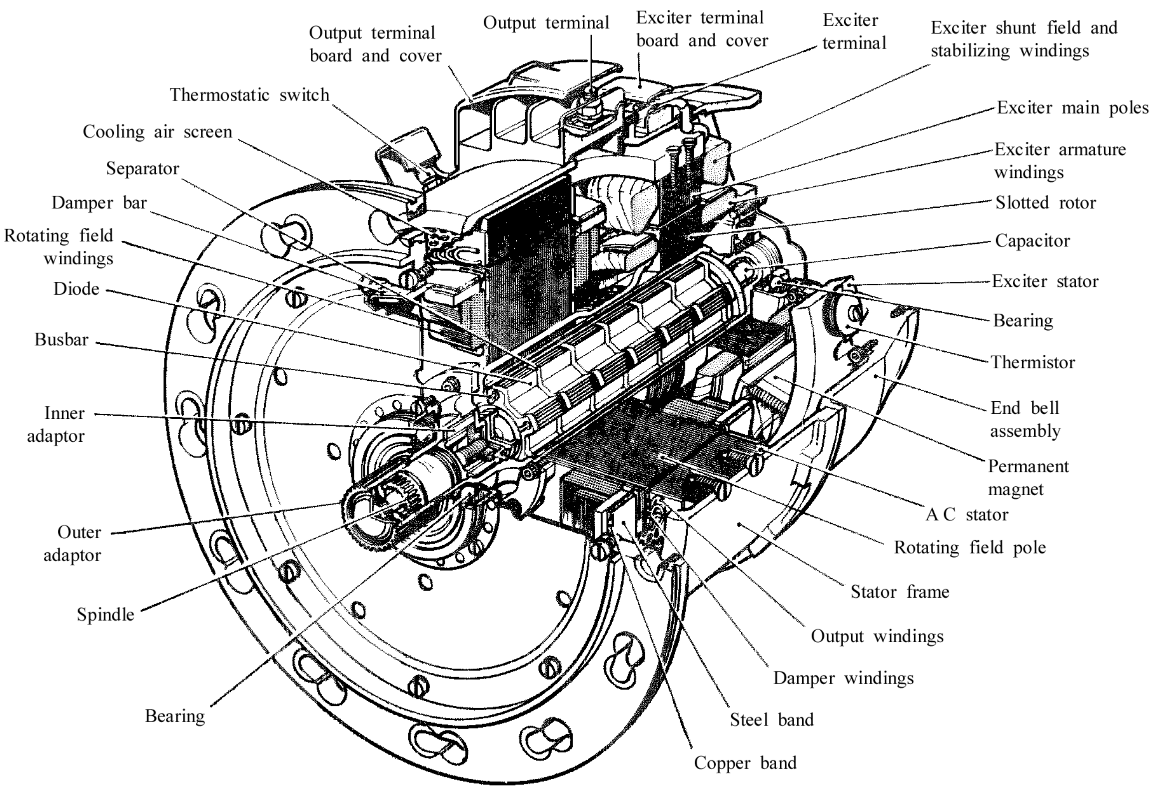



*Fig. 6.4, Constant frequency generator.*

The main generator consists of a three phase star wound stator, and an eight pole rotor and its associated field windings which are connected to the output of the rotating rectifier. The leads from the three stator phases are brought directly to the upper surface of an output terminal board thus preventing the aircraft wiring to the damped directly against the phase leads without current passing through the terminal studs. In addition to the field coils, damper (amortisseur) windings are fitted to the rotor and are located in longitudinal slots in the pole faces large copper bands, under steel bands at each end of the rotor stack, provide the electrical squirrel cage circuit. The purpose of the damper windings is to provide an induction motor effect on the generator whenever sudden changes in load or driving torque tend to cause the rotor speed to vary above or below the normal or synchronous system frequency. In isolated generator operation, the windings serve to reduce excessively high transient voltage caused by line to line system faults, and to decrease voltage imbalance, during unbalanced load conditions. In parallel operation, the windings also reduce transient voltages and assist in pulling in and holding a generator in synchronism.

The drive end of the main rotor shaft consists of a spined outer adaptor which fits over a stub shaft secured to the main generator rotor. The stub shaft, in turn, fits over a drive spindle fixed by a centrally located screw to the hollow section of the shaft containing the rotating rectifier assembly. The complete shaft is supported at each end by pre-greased sealed bearings. The generator is cooled by ram air.

**INTEGRATED DRIVE GENERATORS**

An integrated drive generator is one in which the constant speed drive and generator are mounted side by side to form a single compact unit. This configuration reduces weight, requires less apace, and in comparison with the end to end configuration it reduces vibration. The fundamental construction and operation of both the generator and drive units follow that described above. The essential difference relates to the method of cooling medium, oil is pumped through the generator, the oil itself is in turn cooled by means of heat exchanger system. (Fig. 6.5).

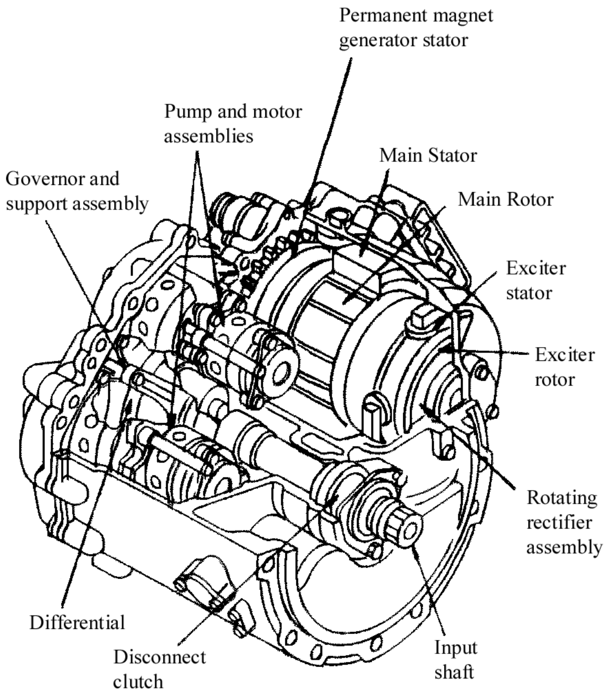



*Fig. 6.5, Integrated Drive Generator.*

**FREQUENCYWILDGENERATORS**

Fig 6.6 is a schematic Dig. of the method adopted for the generator illustrated in Fig 6.2. In this case excitation of the rotor field is provided by d.c. from the aircraft main bus bar. and by rectified a.c. The principle components and the section of the control system, associated with excitation are: the control s/w voltage regulation section field recitation and current compounding section consisting of three phase current transformer and rectifier.

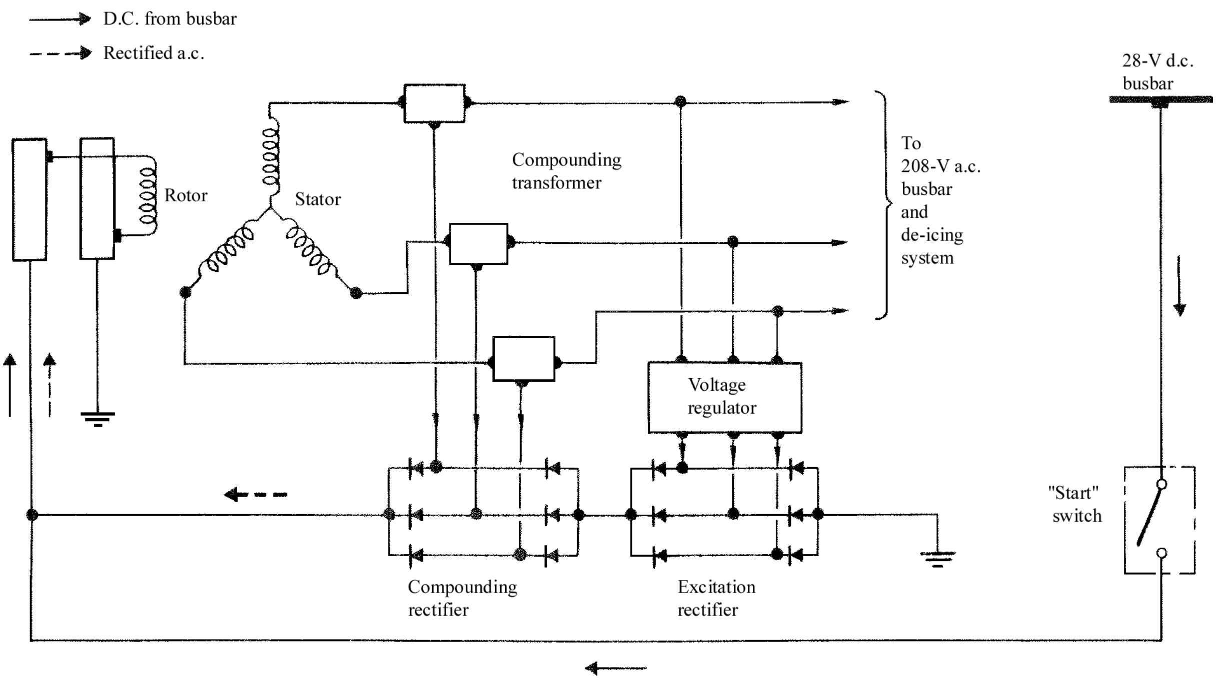



*Fig. 6.6, Frequency-wild generator excitation.*

The primary winding of the compounding transformer are in series with the three phases of the generator and the section winding in the series with the compounding rectifier.

When the control s/w is in the "start" position, d.c. from the main bus bar is supplied to the slip rings, and winding of the generator rotor, thus with the generator running ,a rotating magnetic output is tapped to feed a magnetic field is set up to induce and alternating output in the stator. The output is tapped to feed a magnetic amplifier type of voltage regulator which supplies a sensing current signal to the excitation rectifier. When this signal reaches a predetermine off, load value, the rectifier a.c. through the rotor winding is sufficient for the generator to become self excited and independent of the main bus bar supply which is then disconnected.

The maximum excitation current for wide speed range high output generator is quite high and the variation in excitation current necessary to control the output under varying "load" conditions is such that the action of the voltage regulator must be supplemented by some other medium of variable excitation current .This is provided by the compounding transformer and rectifier and by connecting them in a manner already described direct current proportional to load current is supplied to the rotor field windings

**CONSTANTFREQUENCYGENERATOR**

An excitor stator of the generator is made up of two shunt field winding a stabilizing winding and also six permanent magnetic the later provides a residual magnetic field for initial excitation. A thermistor is located in series with one of the parallel shunt field winding and series as a temperature compensator. At low or normal ambient temperature the high resistance of the thermistor block current flow in its causes the over all shunt field resistance to be about tha of the remaining winding circuit. At the higher temperature resulting from normal operation the resistance of each single circuit increases to approximately double. At the same time ,however the thermistor resistance drops to a negligible value permitting approximately equal current to flow in each winding circuit.

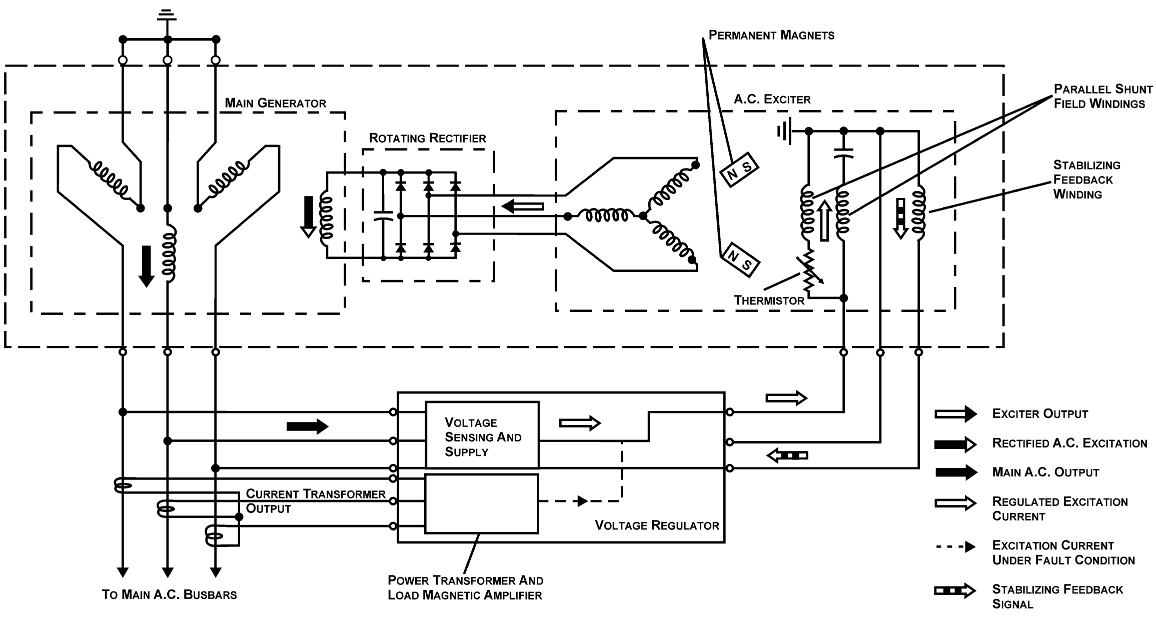



*Fig. 6.7, Circuit diagram of constant frequency generator.*

The stabilizing windings is wound directly over the shunt field windings, and with the permanent magnet poles as a common magnetic core, a transformer type, of coupling between the two winding is thereby provided. The rectified assembly consists of six silicon diodes separated by insulating spacers and connected as the three phase full wave bridge.

The excitation circuit arrangement is shown in figure below when the generator starts running, the flux from the permanent magnet of the a.c exciter provides the initial flow of current in its rotor windings. As a result of the initial current flow, armature reaction is set up and owing to the position of the permanent magnetic poles, the reaction polarizes the main poles of the excitor stator in the proper direction to assist the voltage regulator in taking over excitation control.

The three phase voltage produced in the windings is supplied to the rectifier assembly, the d.c. output of which is in turn, fed to the field coils of the main generator rotor as the required excitation current. A rotating magnetic field is thus produced which induces a three phase voltage put in the main stator windings. The output is tapped and is fed back to the shunt field windings of the excitor, through the voltage regulator system in order to produce a field supplementary to that of the permanent magnets. In this manner the exciter output is increased and the main generator is enabled to build up it output at a faster rate. When the main output reaches the rate value, the supplementary electromagnetic field control the excitation and the effect of the permanent magnets is almost eliminated by the opposing armature reaction. During the initial stages of generator operation, the current flow to the exciter only passes through one of two shunt field windings, due to the increase temperature resistance characteristic of the thermistor. As the temperature of the winding increases, the thermistor resistance decreases to allow approximately equal current to flow in both windings thus maintaining a constant effect of the shunt winding.

In the event that excitation current suddenly increases or decreases as a result of voltage fluctuations due to, for example, switching of loads, a current will be induced in the stabilizing winding since it acts as transformer secondary winding. This current is fed into the voltage regulator as a feed back signal to so adjust the excitation, current, that voltage fluctuations resulting from any cause are opposed and held to a minimum.

■■■

# CHAPTER : 7
# DC MOTORS

### MOTOR PRINCIPLE

When a conductor carrying current is put in a magnetic field, a force is produced on it. The effect of placing a current-carrying conductor in a magnetic field is shown in Fig.7.1. Let us consider one such conductor placed in a slot of armature and suppose that it is acted upon by the magnetic field from a north pole of the motor. By applying left-hand rule it is found the conductor has a tendency to move to the left-hand side. Since the conductor is in a slot on the circumference of the rotor, the force $F_C$ acts in a tangential direction to the rotor. Thus, a torque (turning effect) is developed on the rotor. Similar torques are produced on all the rotor conductors. Since the rotor is free to move, it starts rotating in the anticlockwise direction.

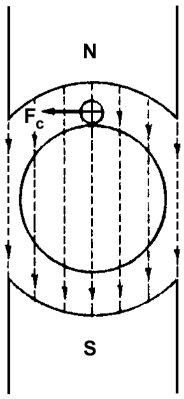



*Fig.7.1. A current-carrying conductor placed in a magnetic field.*

### BACK E.M.F.

When the motor armature rotates, its conductors cut the magnetic flux. Therefore, the e.m.f. of rotation $E_r$ is induced in them. In case of a motor, the e.m.f. of rotation is known as **back e.m.f.** or **counter e.m.f.** The back e.m.f. opposes the applied voltage. Since the back e.m.f. is induced due to generator action its magnitude is, therefore, given by the same expression as that for the generated e.m.f. in a d.c. generator. That is,

$$\boxed{E = \frac{NP\Phi Z}{60\,A}} \qquad (7.3.1)$$

where the symbols have their usual meanings.

### TYPES OF D.C. MOTORS

Direct current motors are named according to the connection of the field winding with the armature. There are three types of d.c. motors.

1. Shunt wound or shunt motor.
2. Series wound or series motor.
3. Compound wound or compound motor.

#### Shunt motor

This is the most common type of d.c. motors. The field winding is connected in parallel with the armature, as shown in Fig. 7.2.

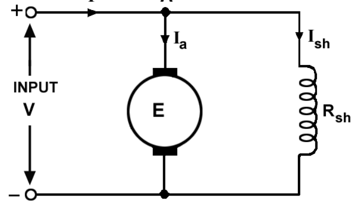



*Fig.7.2. D.C. Shunt motor.*

The current, voltage and power equations for a shunt motor are written as follows :

**Current equation**

By KCL at junction A of Fig. 7.2,

Sum of the incoming currents at A

= sum of the outgoing currents at A

$$I = I_a + I_{sh} \tag{7.6.1}$$

where $I$ = input the current ; $I_a$ = armature current

$I_{sh}$ = shunt field current

**Voltage equations**

The voltage equations are written by using Kirchhoff's voltage law (KVL).

For field-winding circuit

$$V = I_{sh}R_{sh} \tag{7.6.2}$$

For armature-winding circuit

$$V = E + I_aR_a \tag{7.6.3}$$

**Power equations**

Power input = mechanical power developed + losses in the armature + loss in the field

$$VI = P_m + I_a^2R_a + I_{sh}^2R_{sh} \tag{7.6.4}$$

$$= P_m + I_a^2R_a + VI_{sh}$$

$$P_m = VI - VI_{sh} - I_a^2Ra = V(I - I_{sh}) - I_a^2R_a$$

$$= VI_a - I_a^2Ra = (V - I_aR_a)I_a$$

$$\therefore \quad P_m = EI_a \tag{7.6.5}$$

Multiplying Eq. (7.6.3) by $I_a$ we get

$$VI_a = EI_a + I_a^2R_a \tag{7.6.6}$$

$$VI_a = P_m + I_a^2R_a \tag{7.6.7}$$

where $VI_a$ = electrical power supplied to the armature of the motor

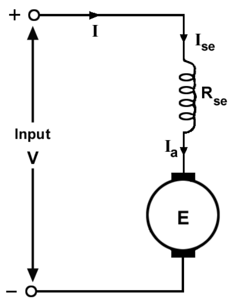



*Fig.7.3, D.C. Series motor.*

**Series motor**

In the series motor (Fig. 7.3), the field winding is connected in series with the armature.

**Current equation**

By KCL in Fig. 7.3

$$I = I_{se} = I_a \tag{7.6.8}$$

where $I_{se}$ = series field current

**Voltage equation**

By KVL in Fig. 7.3

$$V = E + I(R_a + R_{se}) \tag{7.6.9}$$

**Power equations**

Multiplying Eq. (7.6.9) by I we get

$$VI = EI + I^2(R_a + R_{se}) \tag{7.6.10}$$

Power input = mechanical power developed + losses in the armature + losses in the field

$$VI = P_m + I^2R_a + I^2R_{se} \tag{7.6.11}$$

Comparison of Eqs. (7.6.10) and (7.6.11) shows that

$$P_m = EI \tag{7.6.12}$$

**Compound motor**

A d.c. motor having both shunt and series field windings is called a compound motor. It may be a short-shunt compound motor or a long-shunt compound motor. A d.c. compound motor may be cumulatively compounded or differentially compounded as discussed. The current relationships for a compound motor can be written by using KCL. The voltage relationships are written by using KVL.

## CHARACTERISTICS OF A SHUNT OR SEPARATELY EXCITED D.C. MOTOR

In both the cases of shunt and separately excited d.c. motors, the field is supplied from a constant voltage so that the field current is constant. The two motors, therefore, have similar characteristics. Characteristic is a graph between two dependent quantities.

**Speed-armature current characteristics**

In a shunt motor, $I_{sh} = V/R_{sh}$. If V is constant, $I_{sh}$ will also be a constant. Hence the flux is constant at no load. The flux decreases slightly due to armature reaction. If the effect of armature reaction is neglected, the flux $\Phi$ will remain constant. The motor speed is given by

$$N \propto \frac{V - I_a R_a}{\Phi} \tag{7.8.1}$$

If $\Phi$ is constant the speed can be written as

$$N \propto V - I_a R_a \tag{7.8.2}$$

Equation (7.8.2) is the equation of a straight line with a negative slope. That is, the speed N of the motor decreases linearly with the increase in armature current as shown in Fig. 7.4.

Since $I_a R_a$ at full load is very small compared to V, the drop in speed from no load to full load is very small. The decrease in N is partially neutralized by a reduction in $\Phi$ due to armature reaction. Hence for all practical purposes the shunt motor may be taken as a constant-speed motor.

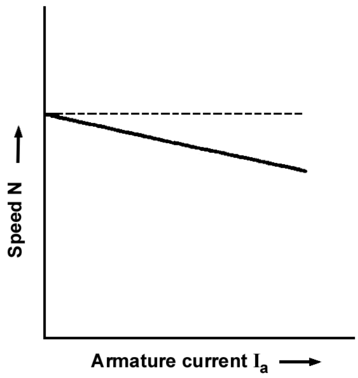



*Fig. 7.4. Speed-armature current ($N/I_a$) characteristic of a shunt or separately excited d.c. motor.*

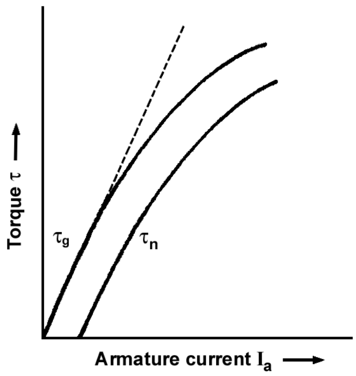



*Fig. 7.5. Torque/armature current characteristic of a shunt or separately excited d.c. motor.*

**Torque/armature current characteristic**

From Eq.

$$\tau_g \propto \Phi I_a \tag{7.8.3}$$

If the effect of armature reaction is neglected, $\Phi$ is nearly constant and

$$\tau_g \propto I_a \tag{7.8.4}$$

Equation (7.8.4) shows that the graph between $\tau_g$ and $I_a$ is a straight line passing through the origin (Fig. 7.5).

If the effect of armature reaction is taken into account, the value of $\Phi$ decreases slightly with the increase in armature current. Hence at higher values of $I_a$ the gross or total torque $\tau_g$ decreases slightly.

The relation between various torques is given by the relation

$$\tau_n = \tau_g - (\tau_f + \tau_w) \tag{7.8.5}$$

where $\tau_n$ = net torque or useful torque or load torque at the output shaft

$\tau_g$ = gross or total torque

$\tau_f$ = frictional torque
$\tau_w$ = windage torque

The graph showing the relationship between the net torque and the armature current is a curve parallel to the corresponding gross torque curve. It is slightly below it.

## CHARACTERISTICS OF A D.C. SERIES MOTOR

### Speed/armature current characteristic

The motor speed N is given by

$$N \propto \frac{V - I_a(R_a + R_{se})}{\Phi} \qquad (7.9.1)$$

At low values of $I_a$, the voltage drop $I_a(R_a + R_{se})$ is negligibly small in comparison with V. Therefore

$$N \propto \frac{V}{\Phi} \qquad (7.9.2)$$

Since V is constant

$$N \propto \frac{1}{\Phi} \qquad (7.9.3)$$

In a series motor, the flux $\Phi$ is produced by the armature current flowing in the field winding so that $\Phi \propto I_a$. Hence the series motor is a variable flux machine. Equation (7.9.3) now becomes

$$N \propto \frac{1}{I_a} \qquad (7.9.4)$$

Thus, for the series motor, the speed is inversely proportional to the armature (load) current. The speed-load characteristic is a rectangular hyperbola as shown in Fig. 7.6.

Equation (7.9.4) shows that when the load current is small, the speed will be very large. Therefore, at no load or at light loads there is a possibility of dangerously high speeds, which may damage the motor due to large centrifugal forces. Hence *a series motor must never run unloaded.* It should always be coupled to a mechanical load either directly or through gearing. It should never be coupled by belt, which may break at any time. With the increase in armature current (which is also the field current) the flux also increases and, therefore, the speed is reduced.

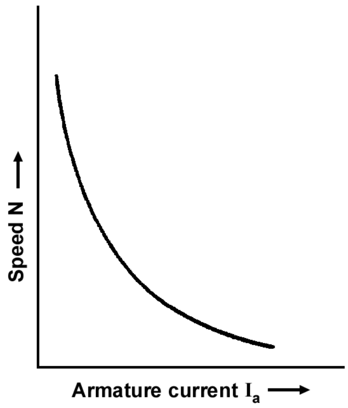



***Fig. 7.6. Speed-armature current characteristic of a d.c. series motor.***

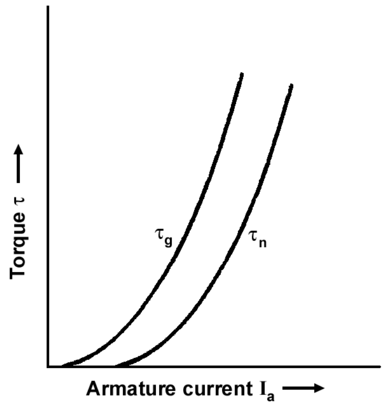



***Fig. 7.7. Torque/armature current characteristic of a d.c. series motor.***

### Torque/armature current characteristic

$$\tau_g \propto \Phi I_a \qquad (7.9.5)$$

Before saturation, $\Phi \propto I_a$ and hence at light loads

$$\tau_g \propto I_a^2 \qquad (7.9.6)$$

Equation (7.9.6) shows that the torque/ armature current ($\tau$/ $I_a$) curve will be parabolic. When the iron becomes magnetically saturated, $\Phi$ becomes almost constant, so that at heavy loads

$$\tau_g \propto I_a \qquad (7.9.7)$$

Equation (7.9.7) shows that the $\tau/I_a$ characteristic is a straight line. Thus, the torque/ current characteristic of a d.c. series motor is initially parabolic and finally becomes linear when the load current becomes large. The characteristic changes smoothly from one curve to another. This characteristic is shown in Fig. 7.7.

The characteristic relating the net torque or useful torque $\tau_n$ to the armature current is parallel to the $T_g/I_a$ characteristic, but is slightly below it (Fig. 7.7). The difference between the two curves is due to friction and windage losses.

**Speed/torque characteristic**

The speed/torque characteristic of a series motor can be derived from its speed-armature current ($N/I_a$) and torque-armature ($\tau/I_a$) characteristics as follows :

For a given value of $I_a$ find $\tau$ from $\tau/I_a$ curve and N from $N/I_a$ curve. This gives one point ($\tau$, N) on speed-torque (N/$\tau$) curve. Repeat this procedure for a number of values of $I_a$ and find the corresponding values of speed and torque $(\tau_1, N_1)$, $(\tau_2, N_2)$ etc. These points are plotted to get the speed-torque characteristic of a d.c. series motor as shown in Fig. 7.8.

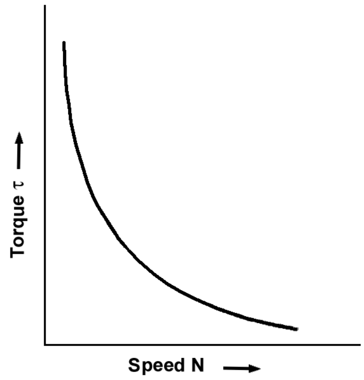


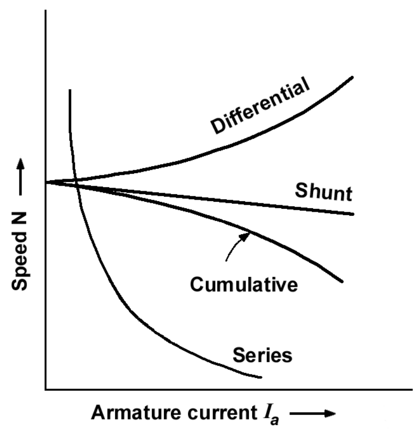

*Fig. 7.8. Speed-torque characteristic of a d.c. series motor.*

This characteristic shows that the d.c. series motor has a high torque at a low speed and a low torque at a high speed. Hence the speed of the d.c. series motor changes considerably with increasing load. It is a very useful characteristic for traction purposes, hoists and lifts where at low speeds a high starting torque is required to accelerate large masses.

**CHARACTERISTICS OF A COMPOUND MOTOR**

A compound motor has both shunt and series field windings, so its characteristics are intermediate between the shunt and series motors. The cumulative compound motor is generally used in practice. The speed-armature current characteristics are shown in Fig. 7.9.

The torque-armature current characteristics are shown in Fig. 7.10.



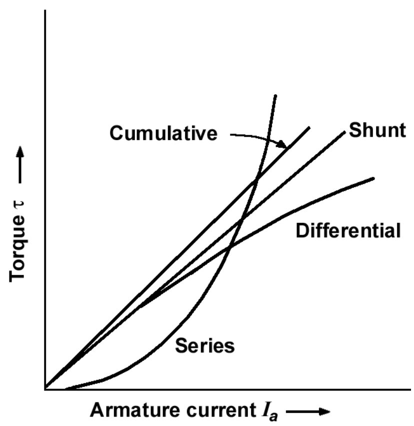




*Fig. 7.9. Speed-armature current characteristic of a d.c. series motor.*

*Fig. 7.10. Torque/armature current characteristic of a d.c. series motor.*

Figure 7.11 shows the speed-torque (N/$\tau$) characteristic of a compound motor. It is found that a compound motor has a high starting torque together with a safe no-load speed. These factors make it suitable for use with heavy intermittent loads such as lifts, hoists etc.

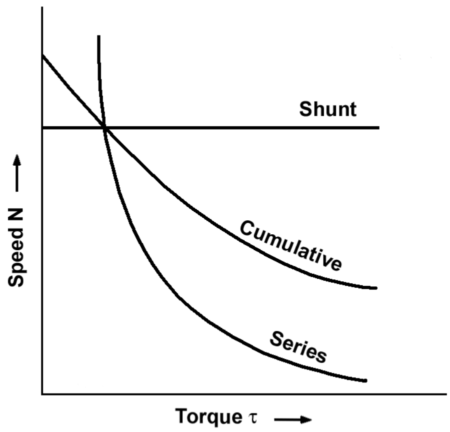



*Fig.7.11. Speed-torque characteristic of a compound motor.*

■■■

# CHAPTER : 8
# AC MOTORS

### CONSTRUCTION

A three-phase induction motor essentially consists of two parts : the stator and the rotor. The stator is the stationary part and the rotor is the rotating part. The stator is built up of high-grade alloy steel laminations to reduce eddy-current losses. The laminations are slotted on the inner periphery and are insulated from each other. These laminations are supported in a stator frame of cast iron or fabricated steel plate. The insulated stator conductors are placed in these slots. The stator conductors are connected to form a three-phase winding. The phase winding may be either star or delta-connected.

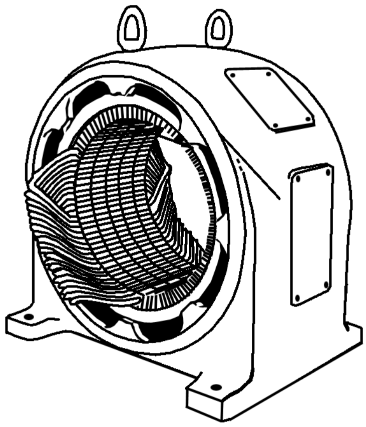


*Fig. 8.1(a), Induction Motor Stator with double-layer winding partly wound.*

The rotor is also built up of thin laminations of the same material as stator. The laminated cylindrical core is mounted directly on the shaft or a spider carried by the shaft. These laminations are slotted on their outer periphery to receive the rotor conductors. There are two types of induction motor rotors :

(a) Squirrel-cage rotor or simply cage rotor.
(b) Phase wound or wound rotor. Motors using this type of rotor are also called slip-ring motors.

### Cage rotor

It consists of a cylindrical laminated core with slots nearly parallel to the shaft axis, or skewed. Each slot contains an uninsulated bar conductor of aluminium or copper. At each end of the rotor, the rotor bar conductors are short-circuited by heavy end rings of the same material. The conductors and the end rings form a cage of the tyre which was once commonly used for keeping squirrels; hence its name. A cage rotor is shown in Fig. 8.1 (b).

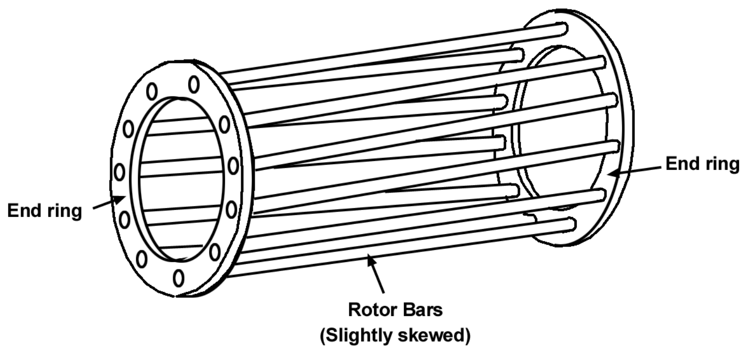



*Fig. 8.1(b), Cage rotor.*

The skewing of cage rotor conductors offers the following advantages :

1. More uniform torque is produced and the noise is reduced during operation.
2. The locking tendency of the rotor is reduced. During locking, the rotor and stator teeth attract each other due to magnetic action.

**Wound rotor or slip ring rotor**

The wound rotor consists of a slotted armature. Insulated conductors are put in the slots and connected to form a three-phase double layer distributed winding similar to the stator winding. The rotor windings are connected in star.

The open ends of the star circuit are brought outside the rotor and connected to three insulated slip rings. The slip rings are mounted on the shaft with brushes resting on them. The brushes are connected to three variable resistors connected in star. The purpose of slip rings and brushes is to provide a means for connecting external resistors in the rotor circuit. The resistors enable the variation of each rotor phase resistance to serve two purposes :

(a) to increase the starting torque and decrease the starting current from the supply.

(b) to control the speed of the motor.

A slip ring induction motor is shown in Fig. 8.2 (a) & (b).

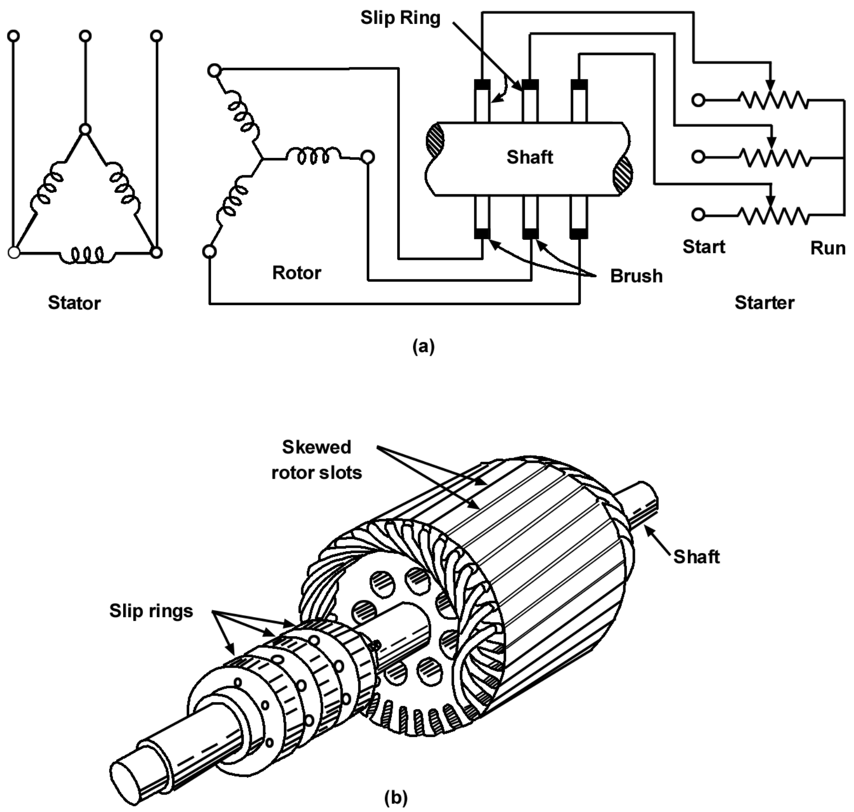



*Fig. 8.2, Slip ring induction motor.*

**COMPARISON OF CAGE AND WOUND ROTORS**

The advantages of the cage rotor are as follows :

(a) Robust construction and cheaper

(b) The absence of brushes reduces the risk of sparking.

(c) Lesser maintenance.

(d) Higher efficiency and higher power factor.

The wound rotors have the following merits :

(a) Higher starting torque and low starting current.

(b) Additional resistance can be connected in the rotor circuit to control speed.

**SINGLE-PHASE INDUCTION MOTOR PRINCIPLE**

A single-phase induction motor consists of a single-phase winding mounted on the stator and a cage winding on the rotor. When a single-phase supply is connected to the stator winding a pulsating magnetic field is produced. By pulsating field we mean that the field builds up in one direction, falls to zero, and then build up in the opposite direction. Under these conditions, the rotor does not rotate due to inertia. Therefore, a single phase induction motor is inherently not self-starting and requires some special starting means. If, however, the single-phase stator winding is excited and the rotor of the motor is started by an auxiliary means, and the starting device is then removed, the motor continues to rotate in the direction in which it is started.

Two theories have been suggested to analyse the performance of a single-phase induction motor, namely the double-revolving-field theory and the cross-field theory. Both the theories are fairly complicated, and neither has any advantage over the other in numerical calculations. Almost similar results are obtained with both the theories. These two theories explain why a torque is produced in the rotor once it is turning. Here we shall discuss the double revolving field theory.

**SPLIT-PHASE INDUCTION MOTOR**

Figure 8.3 shows a split-phase induction motor. It is also called a resistance-start motor. It has a single-cage rotor and its stator has two windings- a main winding and a starting (auxiliary) winding. The main field winding and the starting winding are displaced 90° in space like the windings in a two-phase induction motor. The main winding has very low resistance and high inductive reactance.

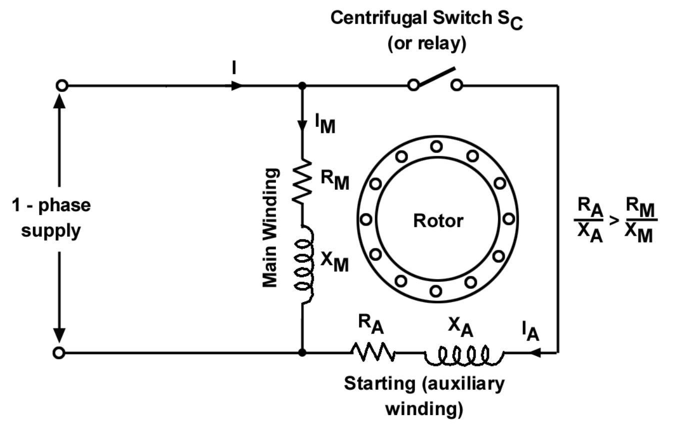
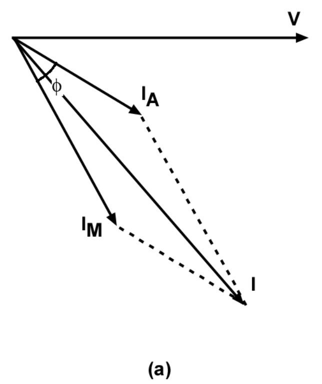



*Fig. 8.3, Split-phase induction motor connections.*

Thus, the current $I_M$ in the main winding lags behind the supply voltage V by nearly 90° [Fig. 8.4 (a)]. The auxiliary winding has a resistor connected in series with it. It has a high resistance and low inductive reactance so that the current $I_A$ in the auxiliary winding is nearly in phase with the line voltage. Thus, there is time phase difference between the currents in the two windings. The time phase difference $\phi$ is not 90° but usually of the order of 30°. This phase difference is enough to produce a rotating magnetic field. Since the currents in the two windings are not equal, the rotating field is not uniform, and the starting torque is small of the order of 1.5 to 2 times the rated running torque. The main and auxiliary windings are connected in parallel during starting. The starting winding are connected in parallel during starting. The starting winding is disconnected from the supply automatically when the motor reaches speed about 70 to 80 per cent of synchronous speed. For motors rated about 100 W or more, a centrifugally operated switch is used to disconnect the starting winding. For smaller motors a relay is often used. The relay is connected in series with the main winding. At the time of starting, a heavy current flows in the relay coil causing its contacts to close. This brings the starting winding into the circuit. As the motor reaches its predetermined speed of the order of 70 to 80 per cent of synchronous speed, the current through the relay coil decreases. Consequently, the relay opens and disconnects the auxiliary winding from the main supply and the motor then runs only on the main windings. The torque-speed characteristic of this motor is shown in Fig. 8.4 (b), which also shows the speed $n_0$ at which the centrifugal operates.



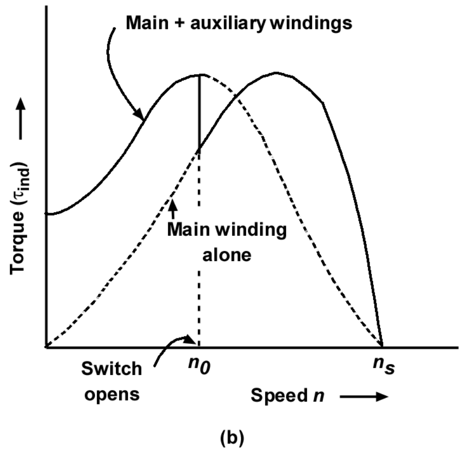




*Fig. 8.4. Resistance split-phase motor (a) Phasor diagram (b) Torque-speed characteristic.*

**Reversal of direction or Rotation**

This motor continues to run in the direction in which it is started. The direction of rotation of the resistance-start induction motor may be reversed by reversing the line connections of either the main winding or the starting winding.

The motor must be brought to rest for this purpose. That is, the reversal of rotation can be made only when the motor is standstill but not while running.

**Motor characteristics**

The starting torque of a resistance-start induction motor is about 1.5 times full-load torque. The maximum or pull-out torque is about 2.5 times full-load torque at about 75 per cent of synchronous speed. The split-phase motor has a high starting current which is usually 7 to 8 times the full-load value.

**Applications**

Split-phase motors are cheap and they are most suitable for easily started loads where frequency of starting is limited. The common applications are washing machines, air-conditioning fans, food mixers, grinders, floor polishers, blowers, centrifugal pumps, small drills, lathes, office machinery, dairy machinery, etc. Because of low starting torques, they are seldom used for drives requiring more than 1 kW.

## CAPACITOR MOTORS

Capacitor motors are single-phase induction motors that employ a capacitor in the auxiliary winding circuit to produce a greater phase difference between the current in the main and auxiliary winnings. There are three types of capacitor motors.

## CAPACITOR-START MOTOR

Figure 8.5 shows the connections of a capacitor-start motor. It has a cage rotor and its stator has two windings namely, the main winding and the auxiliary winding (starting winding). The two windings are displaced 90° in space. A capacitor $C_s$ is connected in series with the starting winding. A centrifugal switch $S_c$ is also connected as shown in Fig. 8.5.

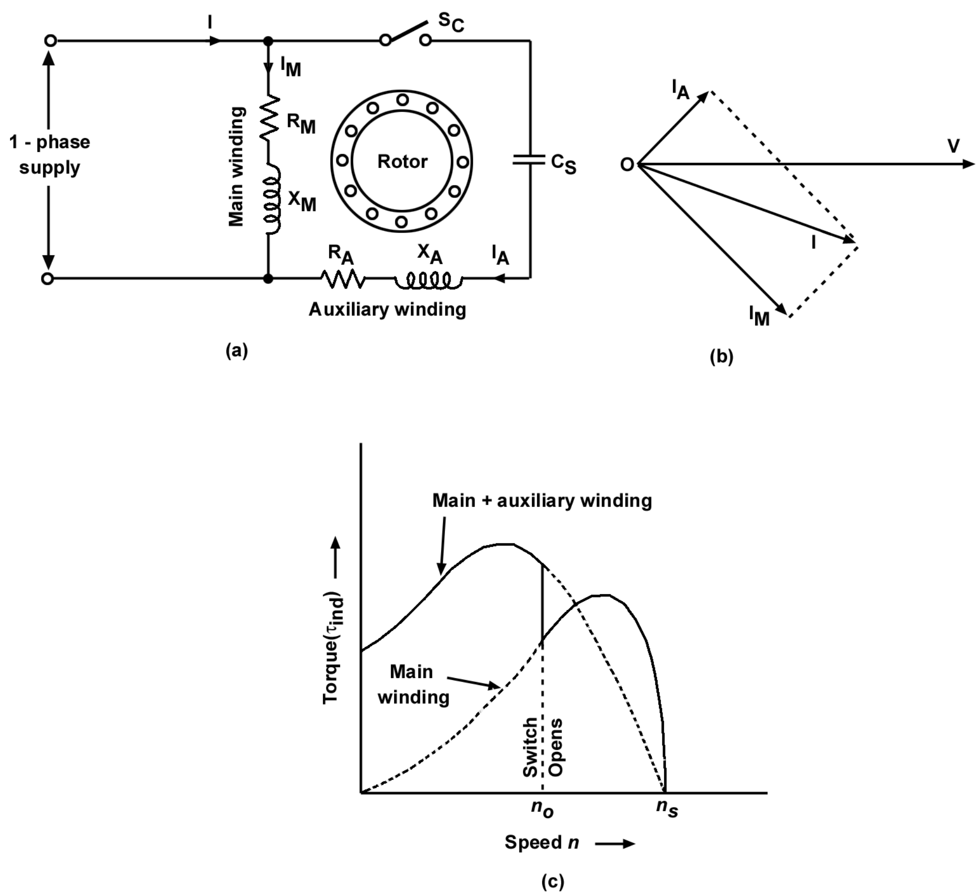



*Fig. 8.5, Capacitor start motor (a) circuit diagram (b) phasor diagram (c) Torque-speed characteristic.*

By choosing a capacitor of the proper rating the current $I_M$ in the main winding may be made to lag the current $I_A$ in the Auxiliary winding by 90°. [Fig. 8.5 (b)]. Thus, a single-phase supply current is split into two phases to be applied to the stator windings. Thus the windings are displaced 90° electrical and their m.m.f.'s are equal in magnitude but 90° apart in time phase.

Therefore the motor acts like a balanced two-phase motor. As the motor approaches its rated speed, the auxiliary winding and the starting capacitor $C_s$ are disconnected automatically by the centrifugal switch $S_c$ mounted on the shaft. The motor is so named because it uses the capacitor only for the purpose of starting.

**Motor Characteristics**

The capacitor-start motor develops a much higher starting torque (3.0 to 4.5 times the full-load torque) than does an equally rated resistance-start motor. The value of the starting capacitor must be large and the starting-winding resistance low to obtain a high starting torque. Because of the high VA rating of the capacitor required, electrolytic capacitors of the order of 250 μF are used. The capacitor $C_s$ is short-time rated. The torque-speed characteristic of the motor is shown in fig. 8.5 (c), which also shows that the starting torque is high.

Capacitor start motors are more costly than split-phase motors because of the additional cost of the capacitor.

**Reversal of Direction of Rotation**

The capacitor-start motor may be reversed by reversing the connections of one of the windings. The motor is first brought to rest for this propose.

**Applications**

Capacitor-start motors are used for loads of higher inertia where frequent starts are required. These motors are most suitable for pumps and compressors, and therefore they are widely used in refrigerators and in air-conditioner compressors. They are also used for conveyors and some machine tools.

**PERMANENT-SPLIT CAPACITOR (PSC) MOTOR**

A permanent-split capacitor (PSC) motor is shown in Fig. 8.6. It has a cage rotor and its stator has two windings, namely, the main winding and the auxiliary winding. This single-phase induction motor has only one capacitor C which is connected in series with the starting winding. The capacitor C is permanently connected in the circuit both at starting and running conditions. A permanent-split capacitor motor is also called the single-value capacitor motor. Since the capacitor C is always in the circuit, this type of motor has no starting switch. The auxiliary winding is always in the circuit, and therefore this motor operates in the same way as a balanced two-phase motor. Consequently, it produces a uniform torque. The motor is therefore less noisy during operations.

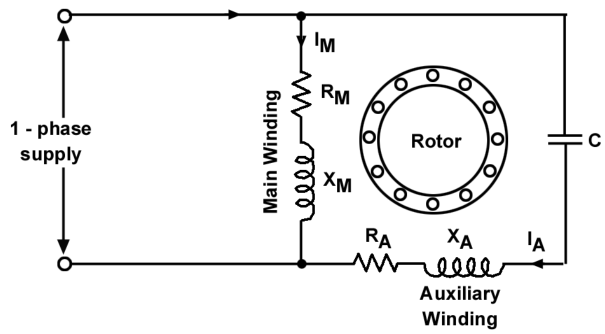



*Fig. 8.6, Permanent-split capacitor motor.*

**Advantages**

a single-value capacitor motor possesses the following advantages :

1. No centrifugal switch is required.
2. It has higher efficiency.
3. It has higher power-factor because of permanently-connected capacitor.
4. It has a higher pull-out torque.

**Limitations**

1. Electrolytic capacitors cannot be used for continuous running. Therefore paper-spaced oil-filled type capacitors are to be used. Paper capacitors of equivalent rating are larger in size and more costly.
2. A single-value capacitor has a low starting torque usually less than full-load torque.

**Applications**

Permanent-split capacitor motors are used for fans and blowers in heaters and air conditioners and to drive refrigerator compressors. They are also used to drive office machinery.

**SHADED-POLE MOTORS**

A shaded-pole motor is simple type of self-starting single-phase induction motor. It consists of a stator and a cage-type rotor. The stator is made up of salient poles. Each pole is slotted on side and a copper ring is fitted on the smaller part a as shown in Fig. 8.7. This part is called the shaded pole. The ring is usually a single-turn coil and is known as shading coil.

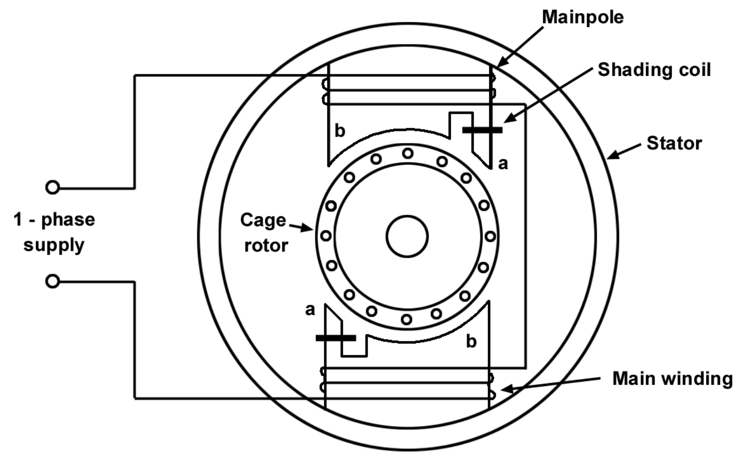



*Fig. 8.7, Shaded-pole motor with two stator poles.*

When alternating current flows in the field winding, an alternating flux is produced in the field core. A portion of this flux links with the shading coil, which behaves as a short-circuited secondary of a transformer. A voltage is induced in the shading coil, and this voltage circulates a current in it. The induced current produces a flux called the induced flux which opposes the main core flux. The shading coil, thus, causes the flux in the shaded portion a to lag behind the flux in the unshaded portion b of the pole. At the same time, the main flux and the shaded pole flux are displaced in space. This space displacement is less than 90°. Since there is time and space displacement between the two fluxes, the conditions for setting up a rotating magnetic field are produced. Under the action of the rotating flux a starting torque is developed on the cage rotor. The direction of this rotating field (flux) is from the unshaded to the shaded portion of the pole. In Fig. 8.7 the direction of rotation is clockwise. In a shaded-pole motor the reversal of direction of rotation is not possible.

**Applications**

Shaded-pole motors are very cheap. The starting torque developed by a shaded-pole motor is very low. The losses are high and the power factor is low. Consequently, the efficiency is also very low. For this reason, the shaded-pole motors are built only in small sizes of the power rating of the order of 40 W or less. They are used to drive devices which require low starting torque. They are most suitable for small devices like relays, fans of all kinds etc. because of their low initial cost and easy starting. The most common applications are table fans, exhaust fans, hair driers, fans for refrigeration and air-conditioning equipments, electronic equipment, cooling fans etc. They are also used in record players, tape recorders, slide projectors, photo copying machines, in starting electric clocks and other single-phase synchronous timing motors.

**COMPARISON BETWEEN SINGLE-PHASE AND THREE-PHASE INDUCTION MOTORS**

Most single-phase induction motors are constructed in fractional kilowatt capacity and are used in places where three phase supply is not readily available. Single-phase motors when compared with 3-phase induction motors have the following disadvantages :

(i) Single-phase motors develop about 50% of the output of that of 3-phase motors for the same size and temperature rise.
(ii) Single-phase motors have lower power factor.
(iii) The starting torque is low in single-phase motors.
(iv) Single-phase motors have lower efficiency.
(v) Single-phase motors are costlier than 3-phase motors of the same rating.

However, single-phase induction motors are simple robust, reliable and less expensive for small ratings. They are used in low-power drives in small industries and domestic and commercial applications. They are generally available upto 1 kW rating.

**SINGLE-PHASE SERIES (UNIVERSAL) MOTOR**

The single-phase series motor is a commutator-type motor. If the polarity of the line terminals of a dc series motor is reversed, the motor will continue to run in the same direction. Thus, it might be expected that a dc series motor would operate on alternating current also. The direction of the torque developed in a dc series motor is determined by both field polarity and the direction of current through the armature ($T \propto \phi i_a$). Let a dc series motor be connected across a single-phase ac supply. Since the same current flows through the field winding and the armature, it follows that ac reversals from positive to negative, or from negative to positive, will simultaneously, affect both the field flux polarity and the current direction through the armature. This means that the direction of the developed torque will remain positive, and rotation will continue in the same direction. The nature of the torque will be pulsating and frequency will be twice the line frequency as shown in Fig. 8.8. Thus, a series motor can run both on dc and ac. Motors that can be used with a single-phase ac source as well as a dc source of supply voltages are called universal motors. However, a series motor which is specifically designed for dc operation suffers from the following drawbacks when it is used on single-phase ac supply :

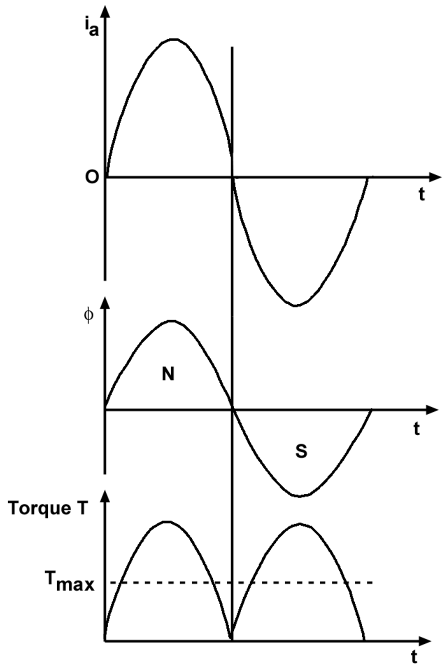



*Fig. 8.8, Developed torque in a single-phase series motor.*

1. Its efficiency is low due to hysteresis and eddy-current losses.
2. The power factor is low due to the large reactance of the field and the armature windings.
3. The sparking at the brushes is excessive.

In order to overcome these difficulties, the following modifications are made in a d.c. series motor that is to operate satisfactorily on alternating current :

(a) The field core is constructed of a material having low hysteresis loss. It is laminated to reduce eddy-current loss.

(b) The field winding is provided with small number of turns. The field pole areas is increased so that the flux density is reduced. This reduces the iron loss and the reactive voltage drop.

(c) The number of armature conductors is increased in order to get the required torque with the low flux.

(d) In order to reduce the effect of armature reaction, thereby improving commutation and reducing armature reactance, a compensating winding is used. This winding is put in the stator slots as shown in Fig. 8.9. The axis of the compensating winding is 90° (electrical) with the main field axis. It may be connected in series with both the armature and field as shown in Fig. 8.10. In such a case the motor is conductively compensated.

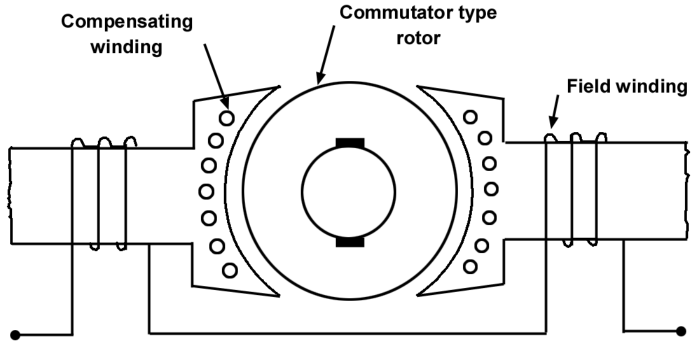



*Fig. 8.9, Series motor with conductively compensated winding.*

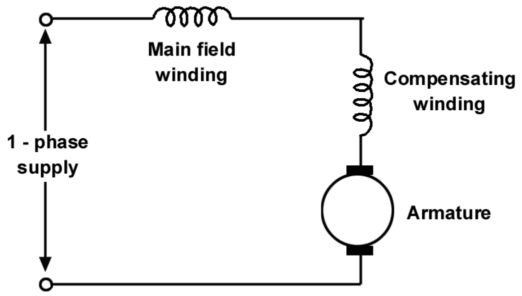



*Fig. 8.10.*

The compensating winding may be short circuited on itself, in which case the motor is said to be inductively compensated (Fig. 8.11).

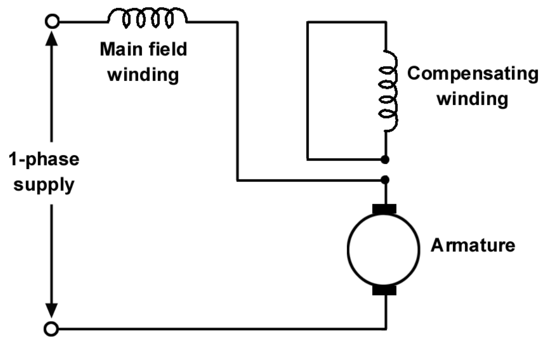



*Fig. 8.11, Series motor with inductively compensated winding.*

The armature of universal motors is of the same construction as ordinary series motor. In order to minimize commutation problems, high resistance brushes with increased brush area are used. The stator core and yoke are laminated to reduce eddy-current loss produced by alternating flux. The machine is generally operated at a lower flux density using very short air gaps.

The universal motor is simple, and cheap. It is used usually for rating not greater than 750 W.

The characteristics of universal motors are very much similar to those of d.c. series motors, but the series motor develops less torque when operating from an a.c. supply than when working from an equivalent d.c. supply. The direction of rotation can be changed by interchanging connections to the field with respect to the armature as in D.C. series motor.

Speed control of universal motors is best obtained by solid-state devices. Since the speed of these motors is not limited by the supply frequency and may be as high as 20,000 r.p.m. (greater than the maximum synchronous speed of 3000 r.p.m. at 50 Hz), they are most suitable for applications requiring high speeds.

There are numerous applications where universal motors are used, such as portable drills, hair dryers, grinders, table-fans, blowers, polishers, kitchen appliances etc. They are also used for many other purposes where speed control and high values of speed are necessary. Universal motors of a given horse power rating are significantly smaller than other kinds of a.c. motors operating at the same frequency.

## SYNCHRONOUS MOTOR

### Construction

The construction of a 3-phase synchronous motor is essentially the same as that of a synchronous generator. The three-phase armature winding is on the stator and is wound for the same number of poles as the rotor. The rotor of a synchronous motor can be of the salient-pole or cylindrical-pole type of construction. Generally, it is of salient-pole

type, except for exceedingly high speed machines. An additional set of windings, called the damper winding, is mounted on the rotor. This winding is placed in slots located in the pole faces and parallel to the shaft as shown in Fig. 8.12. The ends of the copper bars are short-circuited in the same manner as the cage rotor of an induction motor. Damper windings provide a means of starting the synchronous motor. They also serve to increase the stability of the motor during load transients.

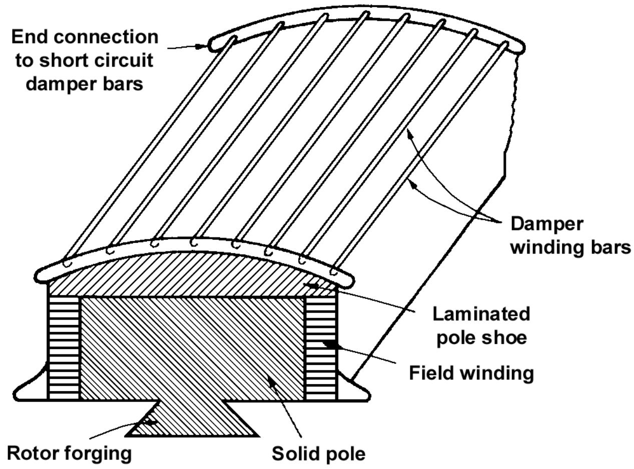



*Fig.8.12. Pole of a synchronous motor showing damper windings.*

A synchronous motor is a doubly excited machine, its armature winding is energized from an a.c. source and its field winding from a d.c. source.

**PRINCIPLE OF OPERATION**

Consider the 2-pole synchronous motor shown in Fig. 8.13. When a three-phase a.c. voltage is applied to the stator winding, a rotating magnetic field is produced in the air gap. The stator field rotates at synchronous speed.

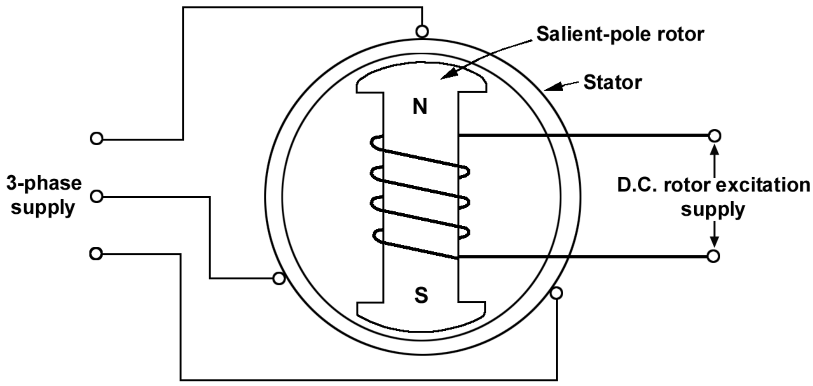



*Fig.8.13. A 2-pole synchronous motor.*

The field current of the motor produces a steady-state magnetic field. Therefore, there are two magnetic fields present in the machine. The rotor will tend to align with the stator field just as two bar magnets will tend to align if placed near each other. Since the stator magnetic field is rotating, the rotor magnetic field and the rotor will tend to rotate with the rotating field of the stator. In order to develop a continuous torque, the two fields must be stationary with respect to each other. This is possible when the rotor also rotates at synchronous speed. The basic principle of synchronous motor operation is that the rotor "chases" the stator magnetic field. In other words, the stator rotating magnetic field tends to "drag" the rotor along, as if north pole on the stator "locks in" with a south pole of the rotor.

Let us assume that the rotor is stationary. When a pair of rotating stator poles sweeps across the stationary rotor poles at synchronous speed, the stator poles will tend to rotate the rotor in one direction and then in the other direction. However, because of the rotor inertia, the stator field slides by so fast that the rotor cannot follow it. Consequently, the rotor does not move and we say that the starting torque is zero. In other words, a synchronous motor is not self-stating.

Let us now assume that the rotor is also rotating at synchronous speed.

**MAIN FEATURES OF SYNCHRONOUS MOTOR**

Some characteristic features of synchronous motor are as follows :

1. It runs either at synchronous speed or not at all. That is, while running it maintains a constant speed. The speed is independent of load.
2. It is not inherently self-starting. It has to be run upto synchronous speed by some means before it can be

synchronized to the supply.

3. It can be operated under wide range of power factors both lagging and leading.
4. It will stall if, while running, the counter torque is increased beyond the maximum torque that the machine can develop.

■■■

# CHAPTER : 9
# VOLTAGE REGULATION

## GENERAL

Voltage regulators are used in aircraft primary power supply systems to maintain the system voltage within the limits necessary for the correct operation of the associated electrical services. In addition, they are, in some cases, used to control the sharing of load between generators operating in parallel.

Depending on the size of the aircraft and design of the generating system, regulators may be of the single unit type operating in conjunction with separate reverse current cut-out relays, voltage differential sensing relays and paralleling relays, or integrated with these components to form special control units or panels.

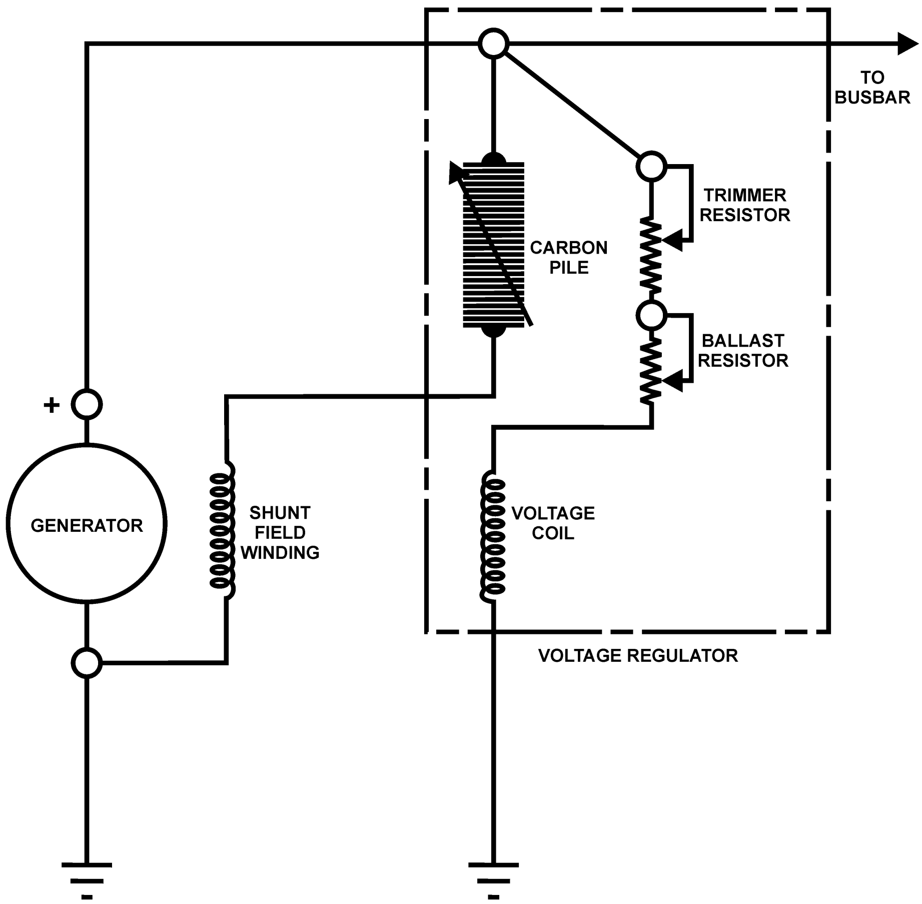



*Fig. 9.1, Carbon-Pile Voltage Regulator Principle.*

## PRINCIPLES

The basic requirement of maintaining a substantially constant voltage in an aircraft power supply system is achieved by the automatic control of the generator field strength, using various types of voltage regulator. The principles of operation of some of these types are contained in the following paragraphs.

### Voltage Regulation

The efficient operation of aircraft electrical equipment requiring d.c. depends on the fundamental requirement that the generator voltage at the distribution busbar system be maintained constant under all conditions of load and at varying speeds, within the limits of a prescribed range. It is necessary, therefore, to provide a device that will regulate the output voltage of a generator at the designed value and within a specified tolerance.

There are a number of factors which, either separately or in combination, affect the output voltage of a d.c. generator, and of these the one which can most conveniently be controlled is the field circuit current, which in its turn controls the flux density. This control can be affected by incorporating a variable resistor in series with the field winding as shown in fig. 9.1. Adjustments to this resistor would vary the resistance of the field winding, and the field current and output voltage would also vary and be brought to the required controlling value. The application of the resistor in the manner indicated is, however, limited since it is essential to incorporate a regulating device which will automatically respond to changes of load and speed, and also, automatically make the necessary adjustments to the generator field current. Three of the regulation methods commonly adopted are : the vibrating contact method ; the one based on the pressure/ resistance characteristics of carbon, namely, the carbon pile method, and the one based on solid-state circuit principles.

***Vibrating Contact Regulator***

Vibrating contact regulators are used in several types of small aircraft employing comparatively low d.c. output generators and a typical circuit for the regulation of both voltage and current of a single generator system is shown in basic form in Fig. 9.4. Although the coil windings of each regulator are interconnected, the circuit arrangement is such that either the voltage regulator only or the current regulator only can operate at any one time. A third unit, called a reverse current cut-out relay, also forms part of some types of regulator, and since the relay has a circuit protection function, a description of its construction and operation will be given.

***Voltage Regulator***

This unit consists of two windings assembled on a common core. The shunt winding consists of many turns of fine gauge wire and is connected in series with the current regulator winding and in parallel with the generator. The series winding, on the other hand, consists of a few turns of heavy gauge wire, and is connected in series with the generator shunt field winding when the contacts of both regulators are closed, i.e. under static condition of the generator system. The contact assembly is comprised of a fixed contact and a movable contact secured to a flexibly-hinged armature. Movement of the armature and, therefore, the point at which contact opening and closing takes place is controlled by a spring which is pre-adjusted to the required voltage setting.

When the generator starts operating, the contacts of both regulators remain closed so that a positive supply can flow through the generator shunt-field winding to provide the necessary excitation for raising the generator output. At the same time current passes through the shunt winding of the voltage regulator and, in conjunction with the series winding, it increases the electromagnetic field of the regulator. As soon as the generator output voltage reaches the pre-adjusted regulator setting, the electromagnetic field becomes strong enough to oppose the tension of the armature spring thereby opening the contacts. In this equilibrium position, the circuit to the series winding is opened causing its field to collapse. At the same time, the supply to the generator field winding passes through a resistance (R) which reduces the excitation current and, therefore, the generator output voltage. The reduced output in turn reduces the magnetic strength of the regulator shunt winding so that spring tension closes the contacts again to restore the generator output voltage to its regulated value and to cause the foregoing operating cycle to be repeated. The frequency of operation depends on the electrical load carried by the generator; a typical range is between 50 to 200 times a second.

In regulators designed for use with twin-generator systems, a third coil is also wound on the electromagnet core for paralleling purposes and is connected to separate paralleling relays.

***Current Regulator***

This unit limits generator current output in exactly the same way as the voltage regulator controls voltage output, i.e. by controlling generator field-excitation current. Its construction differs only by virtue of having a single winding of a few turns of heavy wire.

When electrical load demands are heavy, the voltage output value of the generator may not increase sufficiently to cause the voltage regulator to open its contacts. Consequently, the output will continue to increase until it reaches rated maximum current, this being the value for which the current regulator is set. At this setting, the current flowing through the regulator winding establishes a strong enough electromagnetic field to attract the armature and so open the contacts. Thus, it is the current regulator which now inserts resistance R in the generator shunt-field circuit to reduce generator output. As soon as there is sufficient drop in output the field produces by the regulator winding is overcome by spring tension, the contacts close and the cycle again repeated at a frequency similar to that of the voltage regulator.

**Carbon-pile Voltage Regulator**

The operation of the carbon-pile type of voltage regulator is based on the fact that the contact electrical resistance between faces of carbon discs varies not only with actual area of contact, but also with the pressure by which disc faces are held together. If, therefore, a 'pile' of carbon discs or washers is connected in series with the shunt field winding of a generator the resistance of the field circuit can be varied by adjusting the pressure applied to the 'pile'. (Fig. 10.1)

The necessary variation of pile compression is made through the medium of an electromagnet which opposes the compressive effect of a plate-type control spring. Under static conditions the compressive effect is at a maximum and carbon-pile resistance is at some minimum value. The electromagnet is energized by a voltage coil which is connected across the generator output terminals so that coil current and, consequently, electromagnetic force are substantially proportional to generator output voltage. As the rotational speed of the generator increases, the progressive increase in its voltage results in an increase of electromagnetic force until, at a pre-set voltage level, the electromagnetic force is balanced by the plate-type control spring. If the generator output voltage exceeds the pre-set level, the increase in electromagnetic force overcomes the force of the plate-type control spring and reduces the pile compression, thereby increasing the resistance of the generator shunt field circuit and thus checking the rise in output voltage.

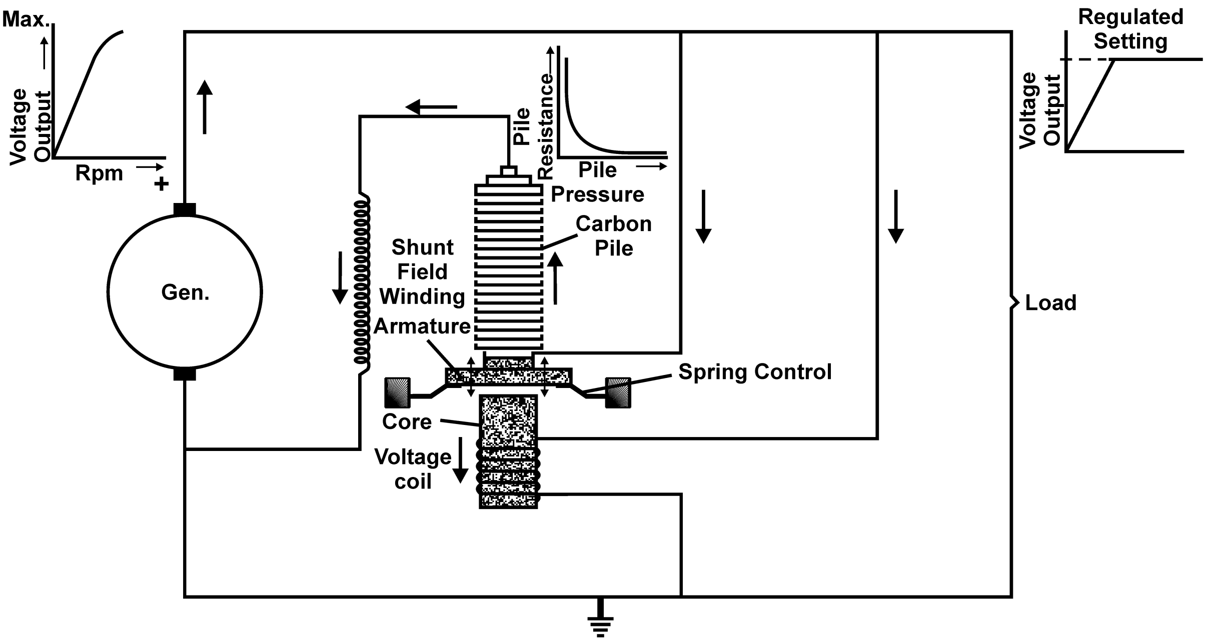



*Fig. 9.2, Carbon-Pile Voltage Regulation.*

**Construction**

The construction of a typical carbon-pile voltage regulator is shown in Figure 9.2. The pile unit is housed within a ceramic tube which, in turn, is enclosed in a solid casing or, more generally, a finned casing for dissipating the heat generated by the carbon pile. Electrical contact at each end of the pile is made by carbon inserts. The initial pressure on the pile is set by a compression screw acting through the pile on the armature and spring plate which is supported on a bi-metal washer. This washer compensates for temperature effects on voltage coil resistance and on any expansion characteristic of the regulator, thus maintaining constant pile compression. The electromagnet assembly comprises a cylindrical yoke in which is housed the voltage coil, a detachable end-plate and an adjustable soft-iron core. The cables from the voltage coil and carbon pile terminate at a connector block or plug on the end plate of the regulator.

**Regulator Adjustments**

Three separate adjustments are normally provided in carbon pile-voltage regulators : (a) voltage coil circuit resistance, (b) magnet core airgap and (c) carbon pile compression.

**(a) Voltage Coil Circuit Resistance**

Adjustment of voltage coil circuit resistance is accomplished by a ballast resistor, pre-set by the manufacturer, to give the correct ampere turns in the voltage coil at the nominal voltage to be controlled. In addition to the ballast resistor, a trimming resistor is also provided for raising or lowering the regulated a voltage level within certain limits, after the regulator is installed in an aircraft.

**(b) Magnet Core Airgap**

The airgap between the magnet core and the armature is pre-set by adjusting the position of the magnet core within the end-plate of the electromagnet housing. The adjustment provides for optimum regulation at the nominal controlled voltage.

**(c) Carbon Pile Compression**

Initial compression of the carbon pile is adjusted by the compression screw to give the correct setting of the plate-type control spring, so that, over the working range of the pile, the spring and magnetic forces exactly counterbalance at any position of the armature. The setting of the screw may be regarded as the characteristic setting of the regulator, and determines the degree of regulation and the stability factor.

**Vibrator-type Voltage Regulator-SH-I**

This type of voltage regulator usually consists of a voltage regulator, a current limiter, and a reverse current cut-out relay housed in the one metal container.

**PRINCIPLE OF OPERATION-SH-II**

The output from the generator armature, entering through terminal G (Fig. 9.3.) passes through the heavy coil of the current limiter and the current coil of the reverse current cut-out relay, then to earth. No current can flow to the battery or the load busbars until the points of the cut-out relay close. As the generator output voltage rises, the magnetic strength of the voltage coil (shunt winding) of the cut-out relay increases enough to close the contacts and put the generator 'on line'. Load current flowing through the cut-out relay current coil now aids the voltage coil in maintaining the contacts closed. The field coils of the generator are excited from the output of the armature, and the earth circuit of the field is completed through the contacts of both the voltage regulator and current limiter. When the voltage rises to its regulated value, the magnetic pull produced by both coils of the voltage regulator opens the regulator contacts. With the regulator contacts open, the earth circuit of the field now includes the resistor, and the field current drops, as will the output voltage. When the contacts open, the accelerator winding circuit is opened and its field collapses completely, rapidly decreasing the flux, so the spring can close the contacts more rapidly than if this winding were not used.

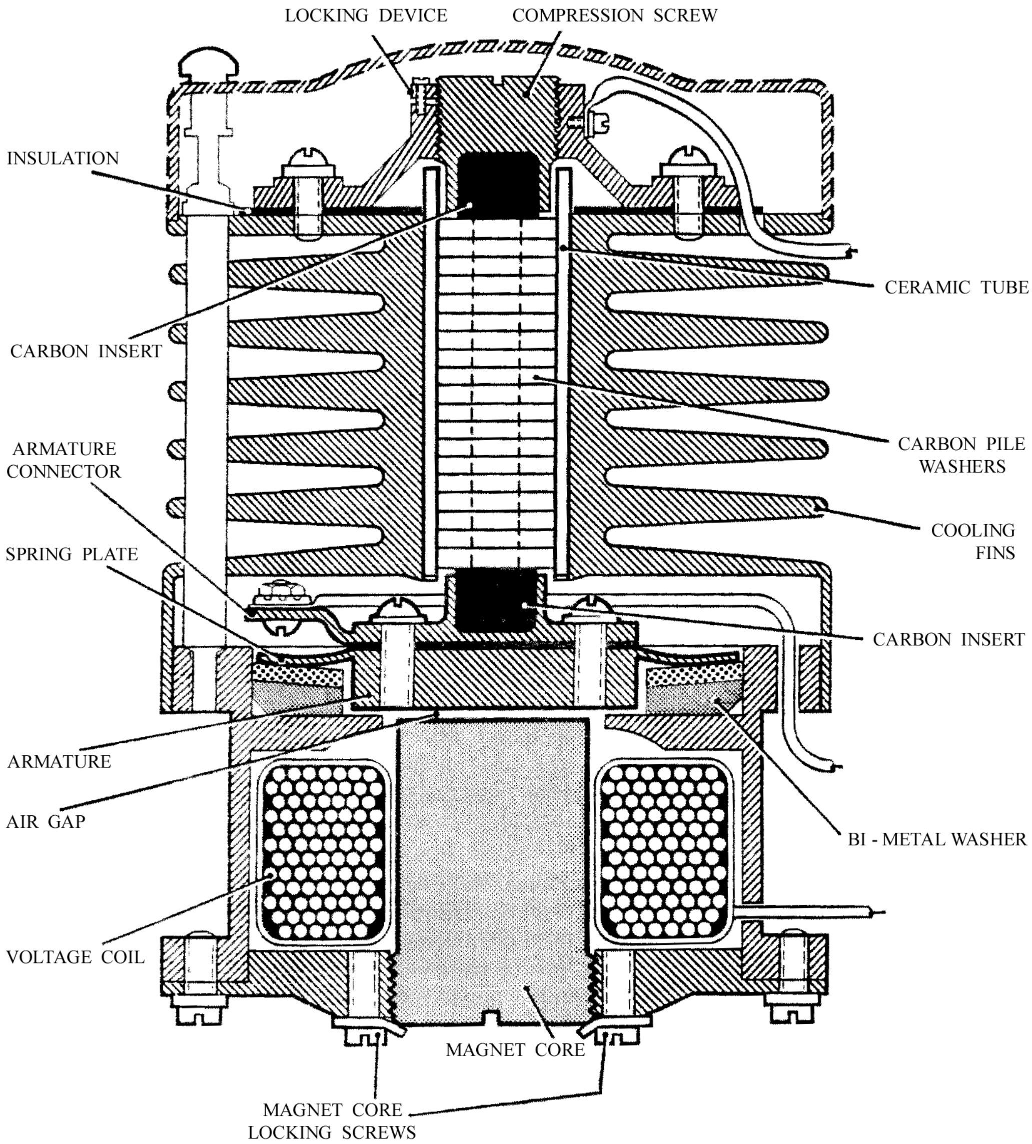



*Fig.9.3, Carbon Pile Voltage Regulator Construction.*

(a) Some regulators employ a second resistor in the regulator housing, in parallel with the generator field. When the field circuit is open-circuited and the field current begins to fall, the inductance of the coil will produce a voltage surge which would tend to cause arcing at the contact points; the purpose of the resistor, therefore, is to suppress the arcing. In normal operation, the contacts open and close between 50 and 200 times per second to maintain the voltage at a constant value.

(b) In some generators one end of the field is connected to the earth brush, and a positive potential is on the field in the regulator to control the voltage. ( Fig 9.5) Before the voltage reaches the regulated level, current flows through the field from the armature of the cut-out relay, through both the voltage regulator and current limiter contacts and through the earth in the generator. When the voltage rises to the pre-set value, the regulator contacts open, and resistor $R_1$ is inserted in the field circuit and reduces the generator output voltage. When the field circuit is opened and the current drops, the induced voltage surge would tend to cause arcing at the contacts, but since resistor $R_2$ is in parallel with the field coils, it will shunt off some of the current and minimize the arcing.

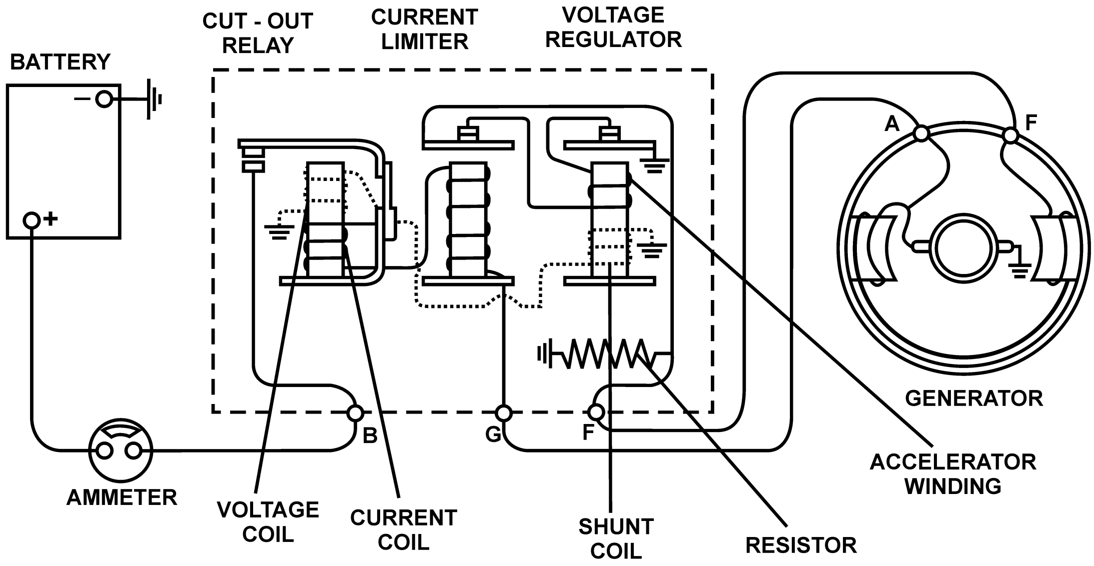



*Fig.9.4, Three-Unit Vibrator Type Voltage Regulator.*

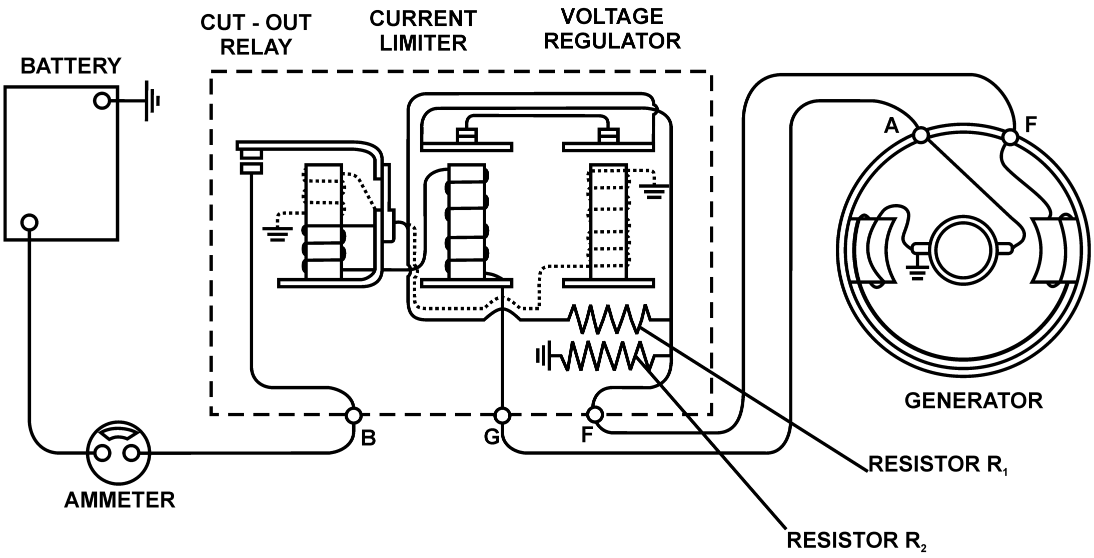



*Fig.9.5, Three Unit Vibrator type Voltage Regulator Connected between the Positive Side of the Generator Armature and the field.*

(c) Some types of vibrator-type voltage regulator have two sets of movable contacts and one fixed contact (fig.9.6). When the engine is running at a relatively slow speed and the field current demands are high, the regulator vibrates between the centre and lower contacts. As the voltage rises to the regulated setting, the contacts open and a resistor is inserted into the field circuit. When the engine speed is high and field current demand is low, the magnetic pull of the voltage regulator is strong enough to cause the contacts to vibrate between the centre and top contacts. (Fig 9.6) When the contacts are open, the resistor is in the field circuit, but when the voltage is high enough to close the top contacts, the field windings are shorted out and no field current flows.

**Current Limiting**

At any time the current drawn by the load reaches the pre-set value, the magnetic field produced by the heavy coil of the current limiter will open the limiter contacts and insert a resistor into the generator field circuit. The increased field resistance will, therefore, lower the output voltage and decrease the current. When the current drops, the contacts close and the voltage again increases. As long as the demands for current exceed the pre-set value, the current-limiter contacts will vibrate.

**Reverse Current Cut-out Relay**

When the generator voltage rises above that of the battery, the magnetic field of the voltage coil in the reverse current cut-out relay will close the contacts and place the generator on line. Load current then flows through the current coil and produces a magnetic field, which aids that produced by the voltage coil and holds the contacts tightly closed. When engine speed decreases and the generator output drops below that of the battery, current will flow from the battery into the generator armature. Current flowing through the current coil of the cut-out relay produces a magnetic field which opposes the field of the voltage coil, and a spring will now open the contacts, taking the generator off line.

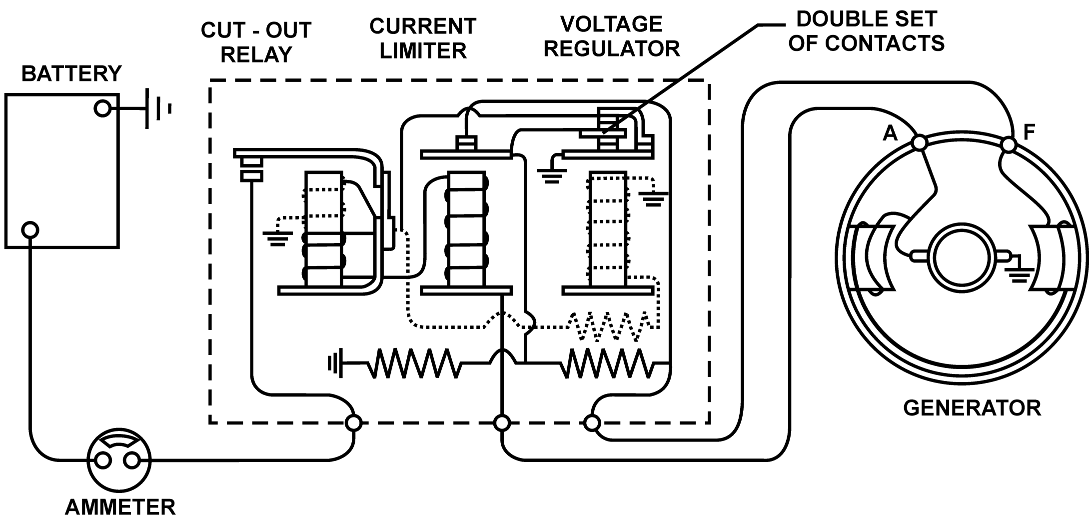



*Fig. 9.6, Three Unit Vibrator Type Voltage Regulator with a Double Set of contacts.*

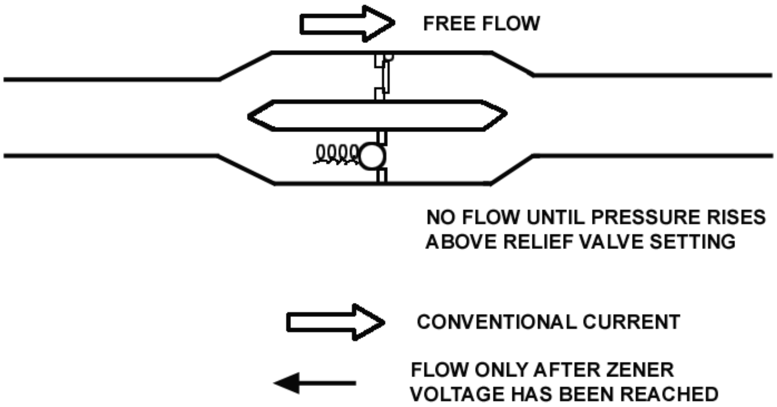



*Fig. 9.7, Principle of a Zener Diode.*

**Solid-state Voltage Regulators**

These generally fall into two categories : transistorized voltage regulators, which use a transistor to actually control the flow of field current but an electromagnetic coil is used to sense the voltage, and transistor voltage regulators, which are fully solid-state and sense the voltage by a zener diode.

## Transistor Voltage Regulators

**(a) Zener Diode**

The principle of operation of a zener diode is that it will allow a free flow of electrons in one direction, but will block the flow in the reverse direction until the voltage has risen to its breakdown, or zener, voltage. This breakdown action does not damage the properties of a zener diode in any way (Fig. 9.7).

**(b) Principle of Operation**

A complete basic circuit of a typical transistor voltage regulator is shown in Figure 9.8. The output of the generator is connected across the voltage divider network of resistors $R_1$, $R_2$ and $R_3$. The zener diode $D_1$ senses the volts drop across $R_1$ and portion of $R_2$. When the voltage across $D_1$ is low, there is no current flow through the base of driver transistor $T_1$; with no base current there will be no emitter-collector current to produce a volts drop across $R_5$. Base current can flow through the output transistor $T_2$ and will conduct, giving a current flow to the generator field. With the field receiving its full field excitation current, the output voltage will rise, and at the regulation level the voltage across the zener diode will cause it to breakdown. With this break-down, base current will now flow in $T_1$, which causes an emitter-collector current to flow through $R_5$. The voltage build-up across $R_5$ brings the base of $T_2$ to the same potential as its emitter and shuts it off, so no field current can flow through $T_2$, causing the generator terminal voltage to fall. Diode $D_2$ provides a constant voltage drop, so the emitter of $T_2$ will be sufficiently below the level of the line voltage, permitting the current through $R_5$ to bring the base voltage of $T_2$ up to that of its emitter so that $T_2$ is shut off. Diode $D_3$ protects the transistors against voltage surges when field current is suddenly cut off. The rapid collapse of the field would induce a voltage high enough to damage the transistors but is prevented from doing so by $D_3$ conducting the voltage to earth. Diode $D_4$ is a transient suppression diode that protects the transistors from any externally-generated voltage surges, while capacitors $C_1$ and $C_2$ smooth out pulsations and cause the regulator to operate smoothly.

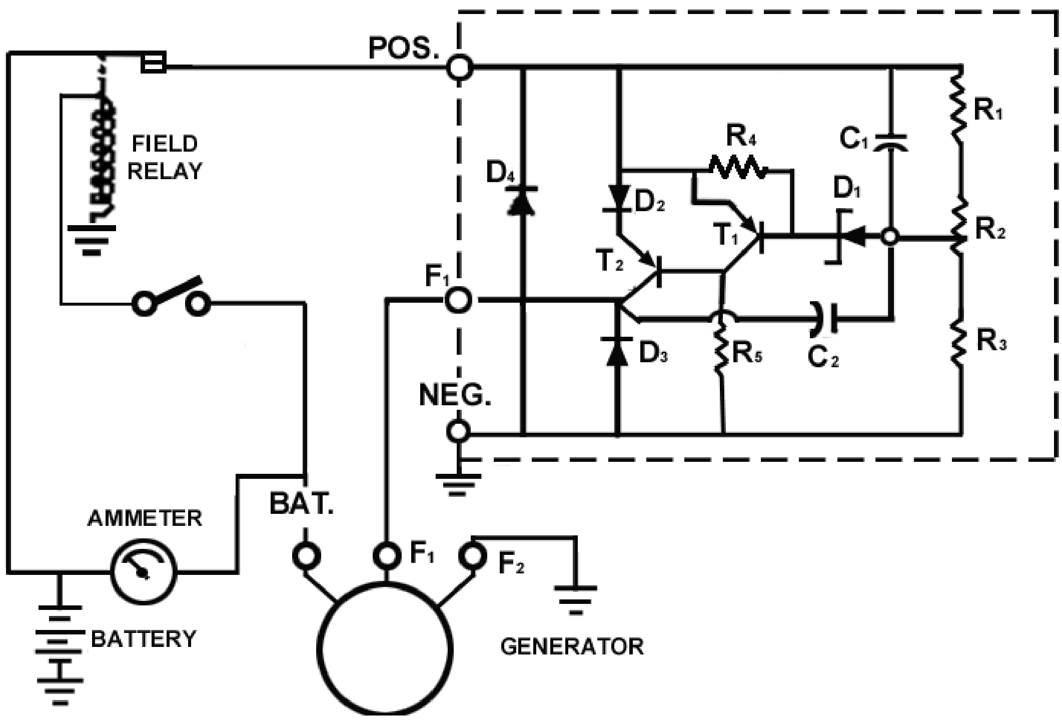



*Fig. 9.8, transistor voltage Regulator.*

■■■

# CHAPTER : 10
# AIRCRAFT BATTERIES

## LEAD-ACID BATTERY

### INTRODUCTION
This chapters gives general guidance on the installation and maintenance of aircraft general purpose lead-acid batteries. The information given should be read in conjunction with the Maintenance Manuals and Overhaul Manuals issued by the battery manufacturers, relevant aircraft Maintenance manuals and approved maintenance Schedules.

### GENERAL CONSTRUCTION
Lead-acid batteries vary in type principally in the thickness of their plates and the type of separators employed. In the conventional type, groups of positive plates, negative plates and separators, are assembled in such a way that the electrolyte solution of sulphuric acid and distilled water can flow freely around the plates. In some types of battery however, the plates and separators are compressed to form a solid block, the separator material being such that it absorbs electrolyte and leaves only a small amount free above the block.

The cell groups are located within containers made of a shock-resistant and acid-resistant material, e.g., polystyrene, and are linked by terminal connecting strips. Vents and plugs are fitted to each cell and are designed to allow gas to escape without leakage of electrolyte. In batteries utilising solid block type cells, the cells are grouped into single blocks (e.g. a 24-volt battery has two single blocks each comprising six individual cells) and are usually enclosed within a polyester bonded fibreglass outer container which also supports the main terminal receptacle.

### CHEMICAL PRINCIPLE
During charging the active material on the positive plates of a cell is converted to lead peroxide and the material on the negative plates to spongy or porous lead. Sulphuric acid is returned to the electrolyte during the charge, gradually strengthening it until the fully charged state is reached. In the charged condition and after a battery has been standing for a specified time (e.g. 8 hours) the open circuit voltage of a cell should be 2.10 to 2.20 volts.

When discharging, both the positive and the negative plates are partially converted to lead sulphate. The sulphuric acid is diluted as part of the process by the formation of water. If lead sulphate in a more permanent form is produced on the plates due to excessive discharge, or other misuse, this will act as a high-resistance component impairing the efficiency and recoverability of the battery.

### MAINTENANCE
The information given in the following paragraphs is intended to serve as a general guide to maintenance practices and precautions to be observed. Precise details concerning a specific type of battery are given in the relevant manuals and approved Maintenance Schedules, and reference must be made to such documents.

**Safety Precautions.**
Lead-acid batteries must be prepared for service, charged, tested, and generally maintained, in a well ventilated workshop area entirely separate from that used for the servicing of nickel-cadmium batteries. This also applies to servicing and test equipment, tools and protective clothing, all of which should be identified as being for use in lead-acid battery servicing only.

Alkaline solutions must not be allowed to come into contact with batteries otherwise severe damage to cell will result.

When handling batteries, or acid, a rubber apron and rubber gloves should be worn; in addition, when dealing with acid, goggles should be worn. After use, these articles should be rinsed free of acid and dried thoroughly. To avoid cracking, or perishing, they should be stored in a cool place, the aprons being hung with as few folds as possible. The gases given off by batteries are highly explosive. Naked lights, therefore, should not be used at any time to examine a battery.

Containers made of a suitable material such as glass, glazed earthenware or ebonite or, alternatively, utensils having a lead lining, should always be used for handling acid or distilled water. When transferring fluids from containers a suitable funnel should be employed.

**NOTE :** Containers filled with distilled water should be stored separately from those containing acid. All containers should be suitably marked, indicating their contents.

When acid has been spilt on the floor of the workshop area or on benches, it should be removed by firstly, washing the affected surfaces with water, then neutralised by washing with sodium bicarbonate solution and lastly, washed again with water.

**NOTE :-** Water and neutralising solution should be soaked up with sawdust, which should afterwards be removed and buried or burned.

If electrolyte comes into contact with the skin, the affected area should firstly be washed with cold water, then neutralised by washing with a sodium bicarbonate solution and lastly, washed with warm water. In the event of electrolyte being splashed into the eyes, they should firstly be washed with cold water, then bathed with 5 per cent solution of sodium bicarbonate, and then again washed with cold water. Immediate medical attention should be obtained in the event of skin burns or eye injury.

**Inspection Before Charging**

All batteries must be inspected before charging and before installation. The following checks are typical of those comprising a battery inspection schedule:-

i) The outside of the battery case should be examined for signs of damage and evidence of locally overheated areas.

ii) The cover, sealing gaskets, or mats as appropriate to the type of battery, should be in good condition.

iii) There should be no evidence of arcing having occurred between the battery and the aircraft structure. If signs of arcing are present, the aircraft battery compartment should be checked to determine whether any insulation provisions have failed, and the necessary remedial action taken. The battery should be cleaned as necessary.

iv) The tops of cells should be inspected for signs of electrolyte leakage, and cleaned and dried where necessary.

v) The battery receptacle should be checked for evidence of burns, cracks and bent or pitted terminals. Defective receptacles should be replaced, because they cause overheating and arcing, and may depress output voltage, which will result in premature battery failure.

vi) All terminals and any exposed cell connecting links must be checked for security, evidence of overheating and corrosion. The terminal nuts, where appropriate, should be tightened to the specified torque values. An acid-free petroleum jelly (e.g. white vaseline) or a silicone base grease, should be lightly smeared onto terminal contacts, connector pins, etc.

**NOTE :** A loose cell link can generate heat and cause arcing which may ignite battery gases.

vii) Vent caps should be checked for security and to ensure that gas exit holes are clear.

Extreme care must be exercised when working around the top of a battery with the cover removed to avoid dropping tools onto the cell connecting links as severe arcing will result, with possible injury to personnel and damage to the battery. Rings, metal watch straps and identification bracelets should not be worn, thereby preventing contact with connecting links and terminals.

**Initial Filling**

A dry uncharged battery must be filled with an electrolyte consisting of battery grade sulphuric acid (see BS 3031) at the relative density recommended by the manufacturer of the particular type of battery, this data being given with the other instructions for filling and charging.

**NOTE :** The relative density of the acid should not be more than 1.300. It is recommended that the acid suppliers be required to lower the relative density to 1.300 [corrected to a temperature of $15^o$ C ($60^o$F)] prior to delivery.

Filling should be carried out methodically to avoid missing any cells. This can be ensured by removing the plug from No. 1 cell and filling as required, then removing the plug from No. 2 cell and fitting it to No. 1 cell,. after which No. 2 cell should be filled and fitted with the plug from No. 3 cell. This procedure should be followed for each cell in numerical order, until the last cell is fitted with the plug from No. 1 cell. On batteries that fill slowly, e.g. those with a solid block plate arrangement which fill by absorption, the vent plugs should be left off unit no more electrolyte is absorbed and the free electrolyte level remains constant.

Although the required electrolyte level will vary with the type and make of battery concerned, in all cases it must cover the top of the plates.

**NOTE :** When poured into the cells, the electrolyte must be at, or only very slightly above, ambient temperature and, if it is obtained by diluting concentrated acid, it must be allowed to cool before use. When diluting concentrated acid, the acid must always be added to the water (at a controlled rate) and never vice versa, since the latter procedure can be extremely dangerous.

After filling, the battery should be allowed to stand for 6 to 8 hours (depending on the manufacturer's instructions for the particular type of battery) so that the battery can cool down, after which the electrolyte level should be restored by adding more electrolyte of the same relative density; the battery is then ready for initial charge.

The relative density of the electrolyte is generally related to a temperature of 15° C (60° F). Readings taken at other temperatures should, therefore, be corrected to 15° C (60° F) as follows :- For the Celsius scale, 0.003 (3 points) should be added to the hydrometer reading for each 4°C by which the temperature of the electrolyte is above 15°C, or 0.003 (3 points) should be subtracted from the hydrometer reading for each 4°C by which the temperature of the electrolyte is below 15°C. Similarly, for the Fahrenheit scale, to correct to 60°F, 0.001 (1 point) should be added or subtracted for every 2.5°F above or below 60°F.

For batteries which are to be used in climates where the temperature frequently exceeds 32°C (90°F) manufacturers sometimes recommend the use of an electrolyte of reduced relative density. For example, a battery may be filled with electrolyte of density 1.260 in temperate conditions, but in tropical conditions the density of the electrolyte may be reduced to 1.230 resulting in a fully charged density of 1.240 to 1.255.

**Charging Conditions**

When charging several batteries, they should be of the same capacity rating, at the same state of discharge and the same recommended charging rate, and connected in series. The number of batteries which may be so connected depends on the voltage available in the charging circuit.

Each group of batteries should be connected to a separate circuit containing an ammeter, voltmeter, variable resistance and other relevant controls.

**NOTE :** Ammeters and voltmeters should be of the moving coil type, or digital presentation type, and checked for accuracy at the periods specified for the charging equipment. Accuracy should be within the values specified for the appropriate type of instrument (see BS 89)

All supply leads and connecting cables must be well insulated, of ample cross sectional area and kept as short as possible. Free ends of cable wires should not be connected to batteries; use should be made of cable end lugs or connector plugs of the type specified for the battery. All connections should be firmly made to give good electrical contact before switching on the charging equipment. To prevent reverse charging, the polarity of the supply leads should be checked with the aid of a centre zero type voltmeter.

It is preferable that neither pole of the charging circuit should be earthed, but if one pole is earthed, it is recommended that the controlling resistance should be between the battery and the unearthed pole.
Batteries requiring different charging rates should not be charged in series, but if this is not possible the limiting current should be that of the battery requiring the lowest charging current.

Vent plugs should be completely unscrewed and lifted, but left in the vent holes before charging is commenced. They should remain in this position during the whole period of charge.

When ready to charge, the variable resistance in the charging circuit should be set in the position of maximum resistance (i.e. minimum current) the charging circuit should be switched on, and the current should be adjusted to the value specified for the particular type of battery.

Charging should, when practicable, be continuous until a fully charged condition is indicated. If charging is interrupted, and batteries are to be left unattended after switching off, both positive and negative supply leads should be disconnected from the batteries.

When cells commence gassing, the voltage and relative density should be measured periodically. In the fully-charged state both values should remain steady.

If a battery in a group should reach the fully-charged condition before the others the charging circuit should be switched off and the charged battery disconnected. The charging current should then be readjusted to a value suitable for continuing the charge of the remaining batteries, and the charging circuit switched on again.

If there is any indication of electrolyte spillage, the affected parts should first be rinsed with water, then with a solution of water and washing soda, and finally sponged with clean water and thoroughly dried. A re-check should be made after 24 hours for any further signs of electrolyte spillage, or corrosion.

**NOTE :** In tropical conditions it is often recommended that the batteries are charged at half the usual rate, with double the charging time.

### Batteries Received Dry and Uncharged

After filling in accordance with specified procedures batteries should be charged, using direct current of correct polarity, at the "initial' charging rate recommended by the battery manufacturer; this will vary from 1 ampere to 5 amperes for a total charging time of about 24 hours.

i) The charge should not be considered complete until the voltage and the relative density (if applicable) of each cell remain constant for the period specified for the type of battery. This period is usually between 3 and 5 hours.

ii) The electrolyte temperature should be checked frequently during the charging period and must not exceed the temperature specified by the battery manufacturer. If the maximum temperature [usually about 60°C (140°F)]is exceeded, the charge should be stopped until the electrolyte temperature has dropped by the specified amount (usually about 12°C or 22°F); or the charging current may be halved and the charging time doubled.

iii) On completion of the charge, any gas in the electrolyte should be released by gently rocking the battery, the electrolyte level then being adjusted as specified by the battery manufacturer.

### Batteries Received with Electrolyte

A battery of the type which forms a solid block with the plates and separators is normally despatched already filled with electrolyte.

i) On receipt of the battery the vent plugs should be removed and the level of the electrolyte checked to ensure that it is approximately 1/4 in above the perforated strip and, if necessary, adjusted to this level using sulphuric acid of relative density 1.270.

ii) The battery should receive a charge current at a value appropriate to the ampere-hour rating of the battery, until the voltage is stable over five consecutive half-hourly readings. Should the temperature of the electrolyte reach 60°C (140°F), the charge should be interrupted until the temperature falls below 43°C(110°F).

iii) During the charge the vent plugs should be kept in, but not screwed down. If the battery was properly filled it should not require any topping up during this time. If the electrolyte level disappears, acid of relative density 1.270 should be added; if the electrolyte level is high the excess should be withdrawn.

### Re-Charging a Battery in Service

The electrolyte level should be checked and, if necessary, adjusted with distilled water, and the battery put on charge at the normal rate recommended. In a fully charged condition, all the cells should gas freely and the relative density of the electrolyte should be within the limits given when corrected for temperature. The main terminal voltage should, under normal temperature conditions, be between 30 and 32.4 volts when measured with the charging current flowing. The charge should be continued until the readings are constant for 3 hours.

**NOTE :** At all times during charging, a check should be kept on the battery temperature to ensure that the maximum permissible limit is not exceeded.

### Electrolyte Level and Adjustments

The periods at which adjustments to the electrolyte should be carried out vary largely with the state of charge and duty cycle of the battery.

The level in the cells must be maintained by the addition of distilled water, as only the water from the electrolyte is lost through electrolysis or evaporation. After initial charge the relative density of the electrolyte should not normally require adjustments, but, in exceptional cases, it may be adjusted in accordance with the manufacturer's instructions.

In order to ensure the mixing of acid with the distilled water, topping-up should be done immediately before a battery is put on charge. Adjustments must be to the correct level to prevent overflow of the electrolyte which could occur as a result of gassing and expansion during the charge. If a battery is to be exposed to very low temperatures, its charge after topping-up, should be prolonged for at least one hour. This will ensure thorough mixing of the electrolyte and thereby avoid the possibility of the water freezing. An additional precaution against freezing is to maintain the battery in a fully-charged state.

For some types of battery, special fillers may be necessary to ensure correct electrolyte level but, in general, the level is specified as a measurement taken from the top of the separator guards or plates.

After adding distilled water, it should be borne in mind that relative density readings cannot be relied upon until the electrolyte and water have been mixed by the gassing of the cells whilst on charge.

A record of the quantity of water added to battery cells should be maintained, since frequent additions are grounds for rejection of cells.

If, one hour after charging, the electrolyte level falls below the specified value, the battery should be reconnected to the charging equipment and the electrolyte level adjusted with the battery on a low charge and slightly gassing.

**State of Charge**

On reaching the fully-charged condition, a lead-acid battery displays three distinct indications: (i) the terminal voltage ceases to rise and remains steady (e.g. 31 volts for a new battery) with charging current flowing, (ii) the specific gravity of the electrolyte ceases to rise and remains constant, and (iii) both sets of plates gas freely. In the conventional type of lead-acid battery all three indications must be in evidence before the battery can be regarded as being completely charged.

**NOTE :** The terminal voltage at the end of charge normally diminishes with the age of a battery. If it is equal to, or below, 28.5 volts a battery should not be put back into service.

In the case of batteries using cells of solid block construction, the state of charge is indicated by the open circuit voltage and gassing. Relative density readings cannot be made since there is insufficient free electrolyte. The method of carrying out an open-circuit voltage check is given in the following paragraphs, the current and voltage values being based on a typical 240 volt 18 ampere-hour (one -hour rate ) battery.

i) Connect the battery to a load that will take approximately 20 amperes, and after current has been flowing for 15 seconds, measure the on-load voltage.
ii) Disconnect the load and take an off-load voltage reading immediately; the increase in reading from on-load to off-load should be approximately 1 volt if the battery is in good condition.
iii) The state of charge is assessed from the off-load voltage. If the voltage is between 25.1 volts and 25.8 volts the battery may be regarded as fully charged. An off-load voltage of between 24.5 volts and 25.1 volts indicates a battery that is from half to quarter discharged. A half-discharged battery will indicate an off-load voltage of between 24.2 to 24.5 volts.

**Capacity tests**

A capacity test should be carried out after initial charge, and thereafter at intervals of three months, or at any time the capacity of a battery is in doubt. Details of test methods are given in the relevant manuals.

The battery should be fully charged and then connected to a suitable discharge control panel incorporating a variable-load resistance, an ammeter and an ampere-hour meter. A separate voltmeter is necessary to measure voltage at the battery terminals, or cell connecting strips.

**NOTE :** If the control panel is not of the automatic type, or if no ampere-hour meter is incorporated, accurate monitoring and control of current must be maintained throughout the tests.

The battery should then be discharged at a rate corresponding to the rating of the battery (as detailed in the appropriate manuals, e.g. an 18 A. H. battery rated at the 1 hour rate would be discharged at 18 amps) until the battery reaches its fully discharged condition. This condition is denoted by the main terminals voltage, or the relative density of the electrolyte, falling to the respective fully discharged values for the particular type of battery. The minimum acceptable capacity for use on aircraft is 80 per cent which, in the case of the example rating quoted, provides a duration of discharge equal to 48 minutes. The result, however, should be compared with previous readings to assess rate of deterioration.

**Insulation Resistance Test**

This test should be carried out at the periods specified in the approved Maintenance Schedule and at any time that electrolyte leakage is suspected.

The battery should be fully charged and the case and cell tops wiped dry. It should then be fitted to a metal base plate by the fixing method normally used in the aircraft. A test should be made between one terminal of the battery and the base plate, using a 250-volt insulation tester, and the minimum insulation resistance obtained must be not less than one megaohm. If a reading below this value is obtained, the battery should be checked for presence of moisture, leaking case, or vented electrolyte, and remedial action taken in accordance with the procedures specified in the relevant manual.

**Leakage Test**

if there is no apparent visible damage to a battery, it should be given a leak test using the tester designed for the particular type of battery. Vent caps should be removed and with the tester held firmly over each vent in turn, a pressure of 14 kN/$m^2$ (2 ibf/$in^2$) should be applied by means of the pump on the test. There should be no detectable leakage after a period of not less than 15 seconds.

**INSTALLATION**

Before installing a battery it should be ensured that it is of the correct type, fully charged and the electrolyte is at the correct level. A capacity test and insulation resistance test must also have been carried out in the manner prescribed for the particular battery. In aircraft using batteries in parallel, it is important to ensure that all batteries are at the same state of charge. Reference should be made to the relevant aircraft maintenance manual for details of the battery system and associated installation instructions. Before coupling the battery connecting plug, a check should be made to ensure that the battery system is switched off and that all electrical services are isolated.

**NOTE :** Batteries are heavy units and require the use of approved and careful handling methods to prevent possible injury to personnel, and damage to the cases or components adjacent to the battery location.

The battery compartment should be clean, dry and free of any acid corrosion or damage. Apart from the mounting tray, which is usually made of an acid-resistant material, the structure adjacent to the battery compartment should be treated with an acid-resistant paint as a protection against the corrosive acid fumes from the battery.

When a battery is located in its compartment, it should be ensured that it is securely attached and that the appropriate clams, or bolts are not over-tightened.

Terminal contacts or connector pins should be lightly coated with an acid-free petroleum jelly (e.g. white vaseline) or a silicone base grease.

The supply cables from the battery should be checked for signs of chafing or other damage; connecting terminals or plugs must be secured without any strain on the terminals, plugs or cables.

Battery installations are normally designed so that in flight, sufficient air is passed through the compartments to dilute the gases given off by the battery, to a safe level. Ventilation systems should, therefore, be checked to ensure there is no obstruction or, if integral venting is used, the system connections should be checked for security and freedom from leaks.

**NOTE :** In some ventilation systems, non-return valves are incorporated in the battery compartment vent lines. These valves should also be checked for security and correct location.

After installation, a check should be made that the electrical connections of the battery supply cables have been correctly made by switching on various electrical services for a specific time period and noting that readings of the aircraft voltmeter remain steady.

**MAINTENANCE OF INSTALLED BATTERIES**

Batteries should be inspected at the periods specified in the approved Maintenance Schedule. The details given in the following paragraphs serve as a guide to the checks normally required.

The battery mounting should be checked for security and the outside of the battery base examined for signs of damage and evidence of local overheating. The latches of the cover should operate smoothly and firmly secure it in position.

Connector lugs or plug pins, should be checked for security and for signs of contamination, burns, cracks, bending or pitting.

Cables should be examined to ensure that their protective covering has not been damaged, and that they have not been affected by dampness or by general climatic conditions prevailing in the battery compartment.
The tops of all cells and vent caps should be inspected for signs of electrolyte leakage and cleaned where necessary.

**NOTE :** When removed, the cover of a battery and cell vent caps should not be placed on any part of the aircraft structure or equipment.

Depending on the type of battery installed, either the relative density, or open-circuit voltage, should be checked to ensure that the values obtained are within the permissible limits.

The electrolyte level should be checked and, if necessary, adjusted with distilled water . The amount of water added to the cells should be recorded. A cell requiring more than the specified amount should be regarded as suspect and the battery should be replaced by a serviceable unit.

**NOTE :-** Batteries should be removed from aircraft in order to carry out electrolyte level adjustments.

The battery ventilation system should be checked to ensure security of connections and freedom from obstruction. Acid drain traps, where fitted, should be checked for signs of acid overflow and, if necessary, removed for cleaning.

During checks on a generator voltage regulator system, it must be ensured that the voltage setting does not cause excessive charging current to be fed to the battery system. Voltage set higher than the specified value, coupled with high ambient temperature, is the most common cause of battery overheating, resulting in a 'thermal runaway' condition and damage to the battery. In some cases consideration may have to be given to aircraft operating in extremely hot climates and the system voltage may have to be reduced. In such cases, the battery may then have to be operated slightly below its maximum capacity.

**NOTE :** Other factors which may cause 'thermal runaway' are inadequate battery ventilation, high relative density of the electrolyte or a low end-of-charge reading.

At the periods specified in the approved Maintenance Schedule, the battery must be removed from the aircraft for capacity and insulation resistance test.

**BATTERY RECORDS**

A technical or service record should be maintained for each battery in service and should provide a fairly comprehensive history of each battery, so that in the event of a malfunction it will assist in determining the problem. The example shown in Figure 1 is intended only as a guide.

**STORAGE AND TRANSPORTATION**

Lead-acid batteries should be stored in a clean, dry, cool, well-ventilated area entirely separate from nickel-cadmium batteries. The area should also be free from corrosive liquids or gases. New batteries may be stored either dry and uncharged, or filled and charged. Batteries of solid block construction may also be stored in the condition in which they are despatched by the manufacturer i.e. filled and uncharged. In this condition only the positive plates are formed so that the batteries remain inert until they are prepared for use. Batteries removed from service must always be stored in the fully-charged condition . The appropriate storage limiting periods must be in accordance with those specified in the relevant manuals. Typical periods are 5 years in a temperate climate for charged or uncharged batteries and from 2 to 3 years in a tropical climate for uncharged batteries, and 18 months for charged batteries. If the storage limiting periods have been exceeded, uncharged batteries should be charged, bench checked or returned to the manufacturer for examination and re-lifing.

Charged batteries should be periodically inspected and given a freshening charge every 2 to 4 weeks. The capacity of batteries should also be checked during the storage period at a frequency which is dictated mainly by their condition. It is recommended that capacity tests be carried out every 6 months for new batteries, and every 3 months for batteries returned from service.

Batteries which have been in use and are discharged, should not be allowed to remain, or be stored in this condition, because of the danger of sulphation of the plates. The lower main terminal voltage limit appropriate to the type of battery should be checked and recharging carried out as necessary; a typical lower limit is 21.6 volts.

If it is necessary to return a battery to the manufacturer, or to an approved overhaul Organisation, it should be prepared in accordance with the transportation requirements specified by the manufacturer for the appropriate battery condition i.e. charged or uncharged. An up-to-date service record should accompany the battery and "This Way Up" international signs affixed to the container.

# LEAD-ACID BATTERY - SERVICE RECORD

MANUFACTURER : ..........................................
DATE OF MANUFACTURE : ...........................
PART NO. & RATING : ......................................
SERIAL NO. : ......................................................

MANUFACTURER : ............................................
DATE OF MANUFACTURE : .............................
PART NO. & RATING : .....................................
SERIAL NO. : .......................................................

| DATE | REASON FOR SERVICE S-Scheduled F-Failure | ELECTROLYTE LEVEL & ADJUSTMENTS (c.c. added) | ELECTRICAL CHECKS | | LEAK TEST | CHARGE | | | REMARKS |
|---|---|---|---|---|---|---|---|---|---|
| | | | CAPACITY | INSULATION RESISTANCE (ohms) | | RATE | RELATIVE | END-OF-CHARGE VOLTS | |
| | | | | | | | | | |

# NICKEL CADMIUM BATTERY

## INTRODUCTION

This chapters gives general guidance on the maintenance and installation of nickel-cadmium batteries (in particular of the semi-open type), which provide a stand-by source of d. c. power in aircraft. It should be read in conjunction with the Maintenance Manuals and Overhaul Manuals issued by the battery manufacturers, relevant aircraft Maintenance manuals and approved maintenance schedules.

## GENERAL DESCRIPTION

Nickel-cadmium batteries may be divided into three ranges of basic design, as described in the following paragraphs.

### Sealed Batteries

This range of batteries consists of those having the cells completely sealed. In general the batteries are of small capacity, and may be used for emergency lighting purposes.

### Semi-sealed Batteries

The cells in this range of batteries are usually mounted in steel containers and are fitted with safety valves. The batteries may be charged fairly rapidly but are very sensitive to overcharge, thus, for aircraft usage, they are usually fitted with a thermal protective device. Under normal conditions the battery requires practically no maintenance beyond periodic cleaning and capacity checks.

### Semi-open Batteries

These batteries are generally used as the main aircraft batteries. The cells are similar in appearance to those of the semi-sealed type, but are deliberately allowed to 'gas' to avoid excessive heating should the battery be on overcharge. The cell cases are usually manufactured from nylon. Because of gassing, the electrolyte has to be topped-up' at periods which vary according to the duty cycle of the battery and the conditions under which it is operated. 'Topping-up' periods are specified in the approved Maintenance Schedule for the aircraft concerned .

## CONSTRUCTION

The plates comprise a sintered base on a nickel-plated steel support. The active materials are nickel hydroxide on the positive plates, and cadmium hydroxide on the negative plates, and these are impregnated into the sintered base by chemical precipitation. This type of plate construction allows the maximum amount of active material to be employed in the electro chemical action.

After impregnation with the active materials, the plates are stamped out to the requisite size. The plates are then sorted into stacks according to the type of cell into which they are to be mounted. Usually there is one additional negative plate for a given number of positive plates. The plates are then welded to connecting pieces carrying the cell terminals, after which a separator is wound between the plates and the insulation is checked under pressure. The plate group is then inserted in the container, the lid secured and pressure-tested for leaks. The separators are usually of the triple layer type, one layer being made from cellophane film, the other two being woven nylon cloth. Cellophane is used because it has low electrical resistivity and is a good barrier material which contributes to the electrical and mechanical separation of the positive and negative plates, and keeps finely divided metal power particles from shorting out the plates while still permitting current flow. It also acts as a gas barrier, preventing oxygen given off at the positive plate during overcharge from passing to the negative plate where it would combine with active cadmium, reduce cell voltage, and produce heat as a result of chemical reaction. The cellophane is prone to damage at high operating temperatures, and failure will result in an adverse change in the operating characteristics of a battery.

The electrolyte is a solution of potassium hydroxide and distilled water, having a relative density of 1.24 to 1.30. It is impregnated into cells under vacuum, after which the cells are given three formation cycles, recharged, and then allowed to stand for a minimum period of 21 days. The discharge characteristics at the end of this period enable the cells to be matched.

In a typical battery each component cell is insulated from the others by its moulded plastic case. All the cells are interconnected via links secured to the terminals of the cells, and are contained as a rigid assembly in the battery case. A vent cap assembly is provided on the top of each cell and, in general, is constructed of plastic, and is fitted with an plastomer sleeve valve. The vent cap can be removed for adjustment of the electrolyte level, and acts as a valve to release gas pressure generated during charging. Except when releasing gas, the vent automatically seals the cell to prevent electrolyte spillage and entry of foreign matter into the cell.

Two venting outlets, a pair of carry-strap shackles, and a two-pin plug for quick-release connection of the aircraft battery system cables, are embodied in the battery case. A removable cover completes the case, and incorporates a pair of slotted lugs which engage with attachment bolts at the battery stowage location.

### Chemical Principle

During charging an exchange of ions takes place' oxygen is removed from the negative plates and is added to the positive plates, bringing them to a higher state of oxidation. These changes continue in both sets of plates for as long as the charging current is applied or until both materials are converted; i.e. all the oxygen is driven out of the negative plates and only metallic cadmium remains, and the positive plates become nickel hydroxide.

The electrolyte acts only as an ionized conductor and is forced out of the plates during charging. It does not react with either set of plates in any way, and its relative density remains almost unchanged. Towards the end of the charging process and during overcharging, gassing occurs as a result of electrolysis which reduces only the water content of the electrolyte. Gassing is dependent on the temperature of the electrolyte and the charging voltage.

During discharge, the chemical action is reversed; the positive plates gradually losing oxygen while the negative plates simultaneously regain lost oxygen. The plates absorb electrolyte to such an extent that it is not visible at the top of the cells.

**MAINTENANCE**
Nickel-cadmium batteries must be prepared for service, charged, tested and otherwise generally maintained, in a well ventilated workshop area which is entirely separate from that used for the servicing of lead-acid batteries. This also applies to servicing and test equipment, tools and protective clothing, all of which should carry some form of identification. Anything associated with lead-acid batteries (acid fumes included) that comes into contact with a nickel cadmium battery or its electrolyte can cause severe damage to this type of battery.

Precise details of inspection and maintenance procedures, and the sequence in which they should be carried out, are given in the relevant battery maintenance and overhaul manuals, and other approved supplementary servicing instructions; reference should, therefore, always be made to such documents. The information given in the following paragraphs is intended to serve as a general guide to the procedures to be carried out appropriate to battery service life and condition, and also to the precautions to be observed.

**Inspection**
The following checks are typical of those comprising a battery inspection schedule:-

a) The battery should be identified to establish any known history. If the battery is a new one a servicing record card should be raised .
b) The outside of the battery case should be examined for evidence of damage, and of locally overheated areas.
c) The battery cover should be removed and its rubber lining inspected for condition. Cover latches should operate smoothly and provide proper security of the cover. Extreme care must be exercised when working around the top of a battery with its cover removed. Tools should not be dropped onto the cell connecting links, as severe arcing will result with possible injury to personnel and damage to the battery. Such personal items as rings, metal watch straps and identification bracelets should be removed, to avoid contact with connecting links and terminals.
d) There should be no evidence of arcing having occurred between the battery and the aircraft structure. The section near the bottom of the case and the slotted lugs of the cover tie-down strap are areas which are most likely to be affected. If signs of arcing are present, the aircraft battery compartment should be inspected and the battery should be completely dismantled and overhauled.
e) The battery should be inspected for signs of electrolyte leakage and should be cleaned where necessary.
f) The battery receptacle should be checked for evidence of burns, cracks and bent or pitted terminals. Defective receptacles, which can overheat, cause arcing and depress output voltage, should be replaced.
g) All cell links should be checked for security and evidence of overheating, and their terminal nuts should be tightened to the specified torque values. Any cell link showing damage to its plating should be replaced.
h) Vent caps should be checked for security and also to ensure that gas exit holes are free from dirt or potassium carbonate crystals. Clogging of vents causes excessive pressures to build up, resulting in cell rupture or distortion of parts. Cell valves, when fitted, should also be checked for security and freedom from dirt or crystal formation. Dirty vent caps or valves should be removed and cleaned.
NOTE: Potassium carbonate is a white crystal formed by the reaction of potassium hydroxide with carbon dioxide in the air; it is non-corrosive, nontoxic, and non-irritating.
j) Temperature sensing devices, when installed, should be checked for secure attachment with leads and connectors showing no signs of chafing or other damage. Electrical checks and/or calibration of these devices should be carried out at the periods specified in the approved Maintenance Schedule.

**Electrolyte Level and Adjustments**
The level of the electrolyte should, depending on manufacturer's recommendations, only be adjusted when a battery is at the end of charge, while still charging, or after a specified standing time. If electrolyte level adjustments were to be made in the discharged or partially discharged condition, then during a charge electrolyte would be expelled from the cells, resulting in corrosive effects on cell links, current leakage paths between cells and battery case, and a reduction in electrolyte density. The manufacturer's instructions regarding checks on electrolyte level and adjustments should be carefully followed and the maintenance kit equipment designed for a particular type of battery should be used.

**NOTE :** Adjustments should not be made when batteries are installed in aircraft.

Only the purest water available, preferably pure de-mineralised or distilled water, should be used for adjusting electrolyte levels, and a record of the quantity added to all cells should be maintained, because it is largely on this evidence that periods between servicing are determined . The 'consumable' volume of electrolyte is normally specified in manufacturer's manuals, but in the absence of such information, a useful guide line is that batteries should not be left for periods which would require the addition of water to any cell by an amount in excess of 1 cc per ampere-hour capacity.

In the event that the electrolyte becomes contaminated, particularly with oil, foaming of the electrolyte will occur. In such cases, a neutralizing fluid, which is available from the relevant battery manufacturer, should be added to the electrolyte, strictly in accordance with the manufacturer's instructions.

Additional potassium hydroxide should not normally be required, but if electrolyte in solution is necessary for topping-up it must be ensured that it is in the proportions specified in the relevant manual.

**NOTE :** Contamination of the electrolyte with tap water, acids, or other non-compatible substances, will result in poor performance or complete failure of a battery.

Potassium hydroxide should be kept in special containers, and because of its caustic nature, should be handled with extreme care to avoid contamination of the person or clothing. Rubber gloves, a rubber apron and protective goggles should always be worn. If contamination does occur, the affected parts should be immediately rinsed with running water. If available, vinegar, lemon juice or a mild boric acid solution may also be used for treatment of the skin. Immediate medical attention is required if the eyes have been contaminated. As a first-aid precaution, they should be bathed with water or a weak boric acid solution, applied with an eye bath.

**Battery Cleaning**
Dirt, potassium carbonate crystals, or other contaminating products, can all contribute towards electrical leakage paths and be a prime cause of unbalanced cells. Cleanliness of batteries is therefore essential.

Deposits should be removed from the tops of cells by using a cloth soaked in de-mineralised or distilled water and stiff fibre bristle brush. Wire brushes or solvents should not be used. If any contaminating product is caked under and around cell connecting links, the links should be removed, if necessary, to facilitate cleaning. Care should always be taken to ensure that debris is not forced down between cells, and in some cases it may be better to scrape deposits loose and then blow them with low-pressure compressed air. The air itself should be clean and dry, and goggles should be worn to protect the eyes.

Some manufacturers specify periodic flushing of cell tops and battery case with de-mineralised or distilled water while brushing away deposits. This method is not recommended, and batteries in a dirty condition, or showing low resistance, should be dismantled and completely serviced.

When it is necessary to clean vent caps and valves, they should be removed from the cells, using the correct extractor tool, and should be washed in warm water to dissolve any potassium carbonate crystals which may have accumulated within the outlet orifices. They should then be rinsed in de-mineralised or distilled water, dried and refitted. Valves should also be tested for correct functioning in accordance with manufacturer's instructions before refitting.

**NOTE :-** Cells should not remain open for longer than is necessary.

**Charging of Batteries**
new nickel-cadmium batteries are normally delivered complete with the correct amount of electrolyte, and in the fully discharged condition. Following a visual check for condition, they must, therefore, be charged in accordance with the manufacturer's instructions before being put into service. Once in service, batteries must then be charged at the periods stated in the approved aircraft Maintenance Schedule. The following information on charging methods and associated aspects is of a general nature only. Precise details are given in relevant manufacturer's manuals and reference must, therefore, always be made to such documents.

**Constant-Current Charging**
This method is the one which should normally be adopted for the workshop charging of batteries, the charging equipment being adjusted and monitored throughout the charging period to supply current either at a single rate, or at several different rates in a stepped sequence. Although more time-consuming than the constant potential method which is often adopted in aircraft battery systems, constant current charging is more effective in maintaining cell balance and capacity. The hour rate of charge current required must be in accordance with that specified by the relevant battery manufacturer.

**NOTE :** The 'hour rate' of a battery refers to the rate of charge and discharge expressed in multiples 'C' amperes, where 'C' is the 1-hour rate. For example, if a battery has a capacity of 23 ampere-hours, then 'C' would be 23 amperes and for a 10-hour rate the charge or discharge current rate would be C/10 amperes i.e. 2-3 amperes.

**Vent Caps**
Before charging, the battery cover should be removed, and with the aid of the special wrench provided in the battery maintenance kit, the vent cap of each cell should also be removed.

**Connection to Charging Equipment**
Charging equipment should not be switched on until after a battery has been connected and the charging circuit has been checked for correct polarity connections.

**Electrolyte Level**
The electrolyte level should be checked and adjusted, as necessary, in accordance with the manufacturer's recommendations.

**Gassing**
Gassing of cells occurs within the region of final charge, as a result of the electrolysis of water into hydrogen and oxygen gases. When gases escape from a cell, the quantity of fluid electrolyte is reduced; vigorous prolonged gassing should therefore be avoided. A "dry" cell is more likely to suffer separator damage, and any cell running hotter than its neighbours should be investigated.

a) The gassing/temperature phenomena provide a useful indication of impending failure of cells; e.g. a cell that gasses sooner and more actively than its neighbours is going to lose more electrolyte, and as a result will run hotter and tend to dry out. Minor differences in gassing are hard to detect, but large differences should not noted and investigated

**State of Charge**
The state of charge cannot be determined by measurement of the electrolyte relative density or battery voltage. Unlike the lead-acid battery, the relative density of the nickel-cadmium battery electrolyte does not change. Except for 'dead' batteries, voltage measurements at either open circuit or on-load conditions do not vary appreciably with state of charge. The only way to determine the state of charge is to carry out a measured discharge test.

**Charging of Individual Cells**
Individual cells must be in an upright position and adequately supported at the sides parallel to the plates during charging. A special frame may be built to fit a cell, or boards or plates may be placed on each side and held together with a clamp. After charging and removal from its support, the sides of a cell should be inspected to ensure there are no bumps or bulges which would indicate an internal failure.

**NOTE :-** Cells should always be fully discharged before removal from a battery and before reassembly.

**Thermal Runaway**
In some small aircraft the battery may be charged by constant potential supplied directly from the d.c. bus-bar. Under correct conditions of temperature and voltage, the internal voltage of the cells rises gradually as the electro-chemical action takes place, and it opposes the charging voltage until this is decreased to a trickle sufficient to balance continuous losses from the cells. The energy supplied to a fully charged battery results in water loss by electrolysis and in heat generation. For a battery in good condition, a point of stability will be reached where heat as a result of trickle current will just balance radiated and conducted heat losses. At low temperatures, a battery will appear to have a limited capacity, and will require more voltage to accept a given amount of charge. As the battery becomes warm, however, its responses return to normal. Operation at high temperatures also limits the capacity, but in such conditions, a battery is subjected to the danger of a 'thermal runaway' condition.

a) At higher than normal temperatures, the heat loss of the battery through radiation and conduction is lower than the heat generating rate and this results in a higher battery temperature. This, in turn, reduces the internal resistance of the battery, so that higher than normal charge current is admitted resulting in an increase in chemical activity, additional heat and a further increase in charging current. This recurring cycle of temperature rise, resistance and voltage drop, and charge current rise, progressively increases the charging rate until sufficient heat is generated to completely destroy a battery.

b) Other factors which can cause overheating of a battery are as follows:-
   i) Voltage regulator of aircraft generating system incorrectly adjusted.
   ii) Frequent or lengthy engine starts at very high discharge rates
   iii) Loose link connections between cells
   iv) Low electrolyte level .
   v) Leakage currents between a cell and battery container and the airframe structure. Periodic measurement of leakage current and removal of any electrolyte that may have accumulated around and between cells should be carried out to prevent high leakage and short circuits from developing
   vi) Use of unregulated, or poorly regulated, ground support equipment to charge a battery, particularly a battery which has become hot as a result of excessive engine cranking or an aborted engine start.
   vii) High initial charging currents imposed on a hot battery.
   viii) Unbalanced cells. Cell unbalance refers to an apparent loss of capacity and to variations in cell voltage at the end of charging cycles. These variations can develop over a period of time, particularly when subjected to operating conditions like those occurring in aircraft utilising charging circuits of the constant potential type.

Other factors which may also contribute to cell unbalance are cell position in the battery, e.g. centre cells run warmer than outer cells, and the self-discharge of individual cells.

c) In some types of aircraft, the batteries specified for use incorporate a thermostat type detector which illuminates a warning light at a preset temperature condition. In addition, a thermistor type sensing network may also be incorporated. The network operates in conjunction with a special solid-state, pulse-charging unit, and its function is to monitor the charging current and to de-energize the charging circuit when the battery temperature exceeds a safe operating limit. Detection devices should be checked at the periods stated in the approved aircraft Maintenance Schedule and in accordance with the relevant manufacturer's instructions.

**Electrical Leakage Check**

Electrical leakage refers to current flowing in a path other than that desired, and in connection with batteries, this means current between the terminals or connectors of cells and any exposed metal on the battery case. The only pertinent measure of leakage of importance to a cell is the rate of discharge caused by the leakage, and this is only significant when its value approaches that specified for the particular type of battery. In one type for example, a leakage of up to 0.020 amps is quoted as the permissible value. Typical methods of determining electrical leakage are described in the following paragraphs.

The positive lead from the terminal of a multi-range test meter should be connected to the positive terminal of the battery and, after selecting the appropriate scale range (usually the one amp. range) the negative terminal lead from the test meter should be touched on any exposed metal of the battery case. If a pointer deflection is obtained it will denote a leakage and the test meter scale setting should be adjusted, if necessary, to obtain an accurate reading which should be within the limits specified. The foregoing check should be repeated between the battery negative terminal and battery case, when again any readings obtained should be within limits. If wither of the readings obtained exceed the specified limits the battery should be thoroughly cleaned and the check again repeated.

If, after thorough cleaning, the leakage current is in excess of the limits it is probable that one of the cells is leaking electrolyte and is, therefore, defective. This cell may be found by measuring the voltage between each cell connecting link and the battery case. The lowest voltage will be indicated at the connecting links on each side of the defective cell which should be replaced.

**Capacity Test**

The capacity or state-of-charge of a fully-charged battery is checked by discharging it at a specified rate (preferably automatically controlled) after it has been standing for a certain time period, and noting the time taken for it to reach a specified on-load voltage. For example, a 23 ampere-hour battery is left to stand for 15 to 24 hours and is then discharged at 23 amperes, i.e. the 1 -hour rate, to 20 volts. A battery should give at least 80% of the capacity specified on its nameplate, or the minimum authorised design capacity, whichever is the greater.

**NOTE :** Some batteries of U. S. origin have initial capacity ratings which are significantly higher than those specified on their nameplates. When the nameplate ratings are no longer obtainable such batteries are rejected.

True capacity must always be recorded, meaning that a full discharge is required, and not one which is terminated when the minimum acceptable level has been reached. Because it is essential to monitor a number of cell voltages very closely, the service of two persons is desirable towards the end of discharge for measurement and recording. At this stage, voltages fall very quickly, and it is highly desirable that measurements be made with a digital voltmeter.

**NOTE :** No cell should be allowed to go into reverse polarity before the measured discharge is complete, and the terminal voltage should not go below 1 volt per cell, since excessive gassing may result.

**Capacity Recycling Procedures**

The purpose of recycling is to restore a battery to its full capability and to prevent premature damage and failure. The discharge rates and voltage values appropriate to the recycling procedures vary between types of battery and reference should always be made to the relevant manual. The figures quoted in the following paragraphs are typical, and serve only as a guide to the limits normally specified.

The battery should be discharged at a current equal to or less than the on-hour rate, and as each cell drops below 0.5 volts (measured by a digital voltmeter) it should be shorted out by means of a shorting strip. The cells should remain in this condition for a minimum period of 16 hours, preferably 24 hours.

**NOTE :** A battery should not be discharged at an excessively high rate and cells then short-circuited, since this produces severe arcing and excessive heat generation.

The shorting strips should then be removed, and the battery charged for 24 hours at the specified recycling charging rate. After approximately five minutes of charge, individual cell voltages should be measured and if any cell voltage is greater than 1.50 volts, distilled water should be added. The amount of water required depends on the rated ampere-hour capacity; a typical maximum value is approximately 1 cc per rated ampere-hour.

After approximately 10 minutes of charge, individual cell voltages should again be measured. Any cell measuring below 1.20 volts or above 1.55 volts should be rejected and replaced.
After 20 hours of charging, individual cell voltages should be measured and recorded, and, if necessary, distilled water should be added to the normal level appropriate to the type of battery.

At the end of the 24 hours charge period, cell voltages should again be measured and compared with those obtained after 20 hours. If the 24 hour voltage reading is below the 20 hours reading by more than 0.04 volts, the cell concerned should be rejected and replaced.

**Cell Balancing**
If a battery fails to give 80% capacity on test, and if premature ageing of some cells is suspected, a cell balancing test should be carried out. The procedure for carrying out the test appropriate to a particular type f battery is prescribed in the relevant manual, and reference should always be made to such document. The following details, based on the test specified for a typical 23 ampere-hour battery, are given only as a general guide.

Note the time, and discharge the battery at 23 amperes until the terminal on-load voltage falls to 20 volts, then stop the discharge. During the discharge, the voltage of each cell should be frequently checked with a digital voltmeter. A zero reading early in the discharge indicates a short circuit cell; a reverse reading indicates a weak cell. In either case the discharge should be stopped, even if the overall battery voltage has not yet fallen to 20 volts. The weak or faulty cell should be shorted out, preferably through a 1 ohm resistor.

Note the time and recommence the discharge at the lower rate of 2.3 amperes. Frequency check the voltage of the cells and short out each cell (with individual shorting strips) as it falls below 1 volt. Record the lapsed time of discharge for the cell to fall below 1 volt, thereby obtaining an indication of the relative efficiency of the cells.

a) Some manufacturers specify 0.5 volts as the point at which shorting of the cells should be carried out. This is satisfactory providing that sufficient time is available to permit shorting of all cells before any are subjected to reverse voltage resulting from the charging effect of stronger cells.

The discharge should be stopped when all the cells are shorted out. The battery should be left in this condition, and also with the main terminals shorted together, for as long as possible, but never less than 16 hours.
The battery should then be charged and the cell-balancing procedure repeated. The discharge times recorded for each cell to fall below 1 volt should show an improvement over those previously recorded.
Weak and internally short-circuited cells should be replaced in accordance with the instructions detailed in the relevant battery Maintenance Manual.

**Voltage Recovery Check**
This check, which should be made at a given time after shorting strips have been removed from the cells or main battery terminals, provides a ready means of detecting high resistance short-circuits and damaged connections within a battery. A typical procedure for this check is given in the following paragraphs.

Shorting strips of one ohm resistance should be connected between cells, and the battery should be allowed to stand for 16 to 17 hours. At the end of this period, the voltage of individual cells should be measured to ensure that they do not exceed the minimum value specified for the battery (a typical minimum value is 0.20 volts).

The shorting strips should then be removed, and after a further standing period of 24 hours, individual cell voltages should again be measured to check their recovery to within normal operating values. A typical minimum value specified as a basis for rejection of a cell is 1.08 volts.

**Insulation Resistance Test**
A test for insulation resistance may be specified by some manufacturers as the means of checking for electrical leakage. Reference should, therefore, be made to the appropriate maintenance manual for the procedure to be adopted, for permissible values, and for any remedial action to be taken.

**Cell Removal and Replacement.**
Cells should be removed from a battery whenever they are suspected of leakage of electrolyte, internal short-circuits, when they fail to balance or if the insulation resistance is found to be below the value specified for the particular battery. The method of removing and replacing cells may vary between types of battery, and the instructions issued by the relevant manufacturers must, therefore, always be carefully followed. The information given in the following paragraphs, although based on a specific type of battery, is intended to serve only as a guide to the practical aspects generally

involved.

The battery should be discharged and the cell links disconnected and removed both from the faulty cell and from the adjoining cells. The cell position should be noted for subsequent entry in the battery record card.

The vent cap should be loosened using the special key provided with the battery maintenance kit.

A cell extractor tool should then be fitted to the cell on the terminals normally used for connecting the cell links. The battery is then held firmly and the cell withdrawn vertically upwards without using undue force. When one cell is removed and all other cell links are disconnected, it is relatively simple to withdraw the remaining cells without the aid of the extractor.

**NOTE :** After removing a cell, its vent cap should be re-tightened.

Cells and the inside of the battery case should be thoroughly cleaned and dried .

After carrying out all necessary checks, serviceable cells should be replaced in the battery case in their correct positions, and a cell-to-cell voltage check should be carried out to ensure that polarities are not reversed. It must be ensured that any new cells are of the same manufacture, part number, and are of matched capacity rating.

**NOTE :** A steady force should be used on terminals to press cells into place. Tight cells should not be hammered into place. For easiest assembly, the cell at the middle of a row should be inserted last.

The surface of cell terminals and connecting links should be clean, and, after ensuring the correct positioning of links, terminal nuts should be tightened to the specified torque value, and in a sequence commencing from the battery positive terminal. Care should always be taken to ensure that nuts actually tighten the connector assemblies, and are not binding as a result of thread damage or bottoming.

**NOTE :** Once a tightening sequence has been started it should be completed, thereby ensuring that a nut has not been overlooked. One loose connection can permanently damage a battery and may cause an explosion.

On completion of cell replacement procedures, the battery should be recharged, tested for insulation resistance, and, if any new cells have been fitted, a capacity test should also be carried out.

**Rejected Batteries or Cells**

Any batteries or cells which are rejected should be conspicuously and permanently marked on their cases to indicate tat they are to be used only for general ground use.

**INSTALLATION**

It should be ensured that the battery is of the correct ampere hour rating, fully charged, and that the electrolyte is at the correct level. Depending on the service history of the battery, appropriate tests, e.g. capacity test, capacity recycling and cell balancing, must also have been carried out in the manner prescribed for the particular battery. Reference should be made to the relevant aircraft Maintenance Manual for details of the battery system and associated installation instructions. Before coupling the system connecting plug, a check should be made to ensure that the battery system switch is OFF, and that all electrical services are isolated.

**NOTE :** Batteries are heavy units, and they require the use of approved handling methods to prevent possible injury to personnel and damage to the cases or components adjacent to the battery location. Vent pipes should not be used for lifting purposes.

The battery compartment should be thoroughly clean and dry, and the battery should be securely attached in its mounting. Clamp nuts should not be over-tightened since distortion of the battery cover may result, which could affect the venting arrangements.

**NOTE :** If a battery compartment has been previously used for lead-acid batteries, it should be washed out with an acid neutralising agent, dried thoroughly, and painted with an alkaline-resistant paint.

The supply cables from the battery, and , where appropriate, thermostat and battery charging system cables, should be checked for signs of chafing or other damage. Cable connecting plugs should be securely made, without any strain on the plugs or cables.

Battery installations are normally designed so that in flight, sufficient air is passed through the compartment to dilute the hydrogen gas given off by a battery, to a safe level. Ventilation systems should therefore be checked to ensure there is no obstruction or, if integral venting is used, the connections should be checked for security and leaks.

**NOTE :** In some ventilation systems, non-return valves are incorporated in the battery compartment vent lines. These valves should also be checked for security and correct location.

After installation, a check should be made that the electrical connections of the battery supply cables have been correctly secured by switching on some electrical services for a specific time period and nothing that readings of the aircraft voltmeter remain steady. A typical load and time is 30 amperes for 30 seconds. For battery systems having a separate 'in-situ' charging the unit should be switched on and its electrical settings checked to ensure proper charging of the battery.

**MAINTENANCE OF INSTALLED BATTERIES**

Batteries should be inspected at the periods specified in the approved aircraft Maintenance Schedule. The details given in the following paragraphs serve as a general guide to the checks normally required.

The battery mounting should be checked for security, and the outside of the battery case should be examined for signs of damage and for evidence of locally overheated areas. The latches of the cover should operate smoothly and should firmly secure the cover in position.

Connecting plugs of the battery receptacle, thermostat and battery charger units, where fitted, should be checked for signs of contamination, burns, cracks, and bent or pitted terminal fittings.

The tops of all cells and vent caps should be inspected for signs of electrolyte leakages and should be cleaned where necessary.

The electrolyte level should be checked, and if any adjustments are necessary, these should be made after removing the battery from the aircraft and checking that it is in the fully charged condition. The amount of water added to the cells should be noted on the battery record card. A cell requiring more than the specified amount should be regarded as suspect, and the battery should be replaced by a serviceable unit. In aircraft having an independent charging unit, the unit should be switched on and the battery charged in accordance with the procedure specified in the relevant aircraft Maintenance Manual.

NOTE: When removed, the battery cover and cell vent caps should not be placed on any part of the aircraft structure or equipment.

The battery ventilation system should be checked to ensure security of connection, and freedom from obstruction.

**BATTERY RECORDS**

A technical or service record should be maintained on each battery in service. Discretion maybe exercised as to the layout of such a record and the extent of the details it should contain. it should, however, provide a fairly comprehensive history of the specific battery, so that in the event of a malfunction it will assist in establishing the fault. The example shown in Figure 1 is intended only as a guide.

**STORAGE AND TRANSPORTATION**

Nickel-cadmium batteries should be stored in a clean, dry, well-ventilated area and should be completely segregated from lead-acid batteries. The area should also be free from corrosive liquids or gases. It is recommended that they should be stored in the condition in which they are normally received from the manufacturer, i.e. filled with electrolyte, discharged and with shorting strips fitted across receptacle pins. Cell connecting strips and terminals should be given a coating of acid-free petroleum jelly (e.g. white vaseline)

The temperatures at which batteries may be stored are quoted in the relevant manuals, and reference should, therefore, be made to these. In general, a temperature of 20°C is recommended for long-term storage.

If batteries are to be stored in a charged condition, they must be trickle charged periodically in order to balance the inherent self-discharge characteristic. Since this discharge is temperature sensitive, the trickle charge rate is, therefore, dependent on the storage temperature conditions.

If it is necessary to return a battery to the manufacturer or to an approved overhaul organisation, it should be discharged, but not drained of electrolyte. It should be packed in its original container, together with its service record and 'This Way Up' international signs affixed to the outside.

**NOTE**: If transportation is to be by air, the container must comply with IATA regulations concerning the carriage of batteries containing alkaline electrolyte.

# NICKEL-CADMIUM BATTERY SERVICE RECORD

## BATTERY AND AIRCRAFT DATA

Manufacturer ..................................

Part No. ...........................................

Serial No. ........................................

Rating : Volts .............. Ah................

Mod. State .......................................

Aircraft Type .....................................

Registration ......................................

Battery Function (e.g. Standby. A.P.U. Starting)

...........................................................

Date Installed ....................................

Hours Flown ......................................

## SERVICING DATA

Date Removed .................................

Date Serviced ..................................

Workshop Ambient Temp. ................

Reason for Removal ..........................

Servicing Instruction Used .................

Date Released ..................................

| **Operation** | **Results/Comments** | **Initials** | |
|---|---|---|---|
| | | **Mech.** | **Insp.** |
| Details of operations performed and measurements required - | | | |
| | | | |

## CELL DATA

| **Position In Battery** | **Serial No.** | **Water Added (c.c.)** | **Voltage** | **Temperature** | **Final Voltage** | **Capacity (Ah)** |
|---|---|---|---|---|---|---|
| *1* | | | | | | |
| *2* | | | | | | |
| *19* | | | | | | |
| *20* | | | | | | |

**MAIN TERMINAL VOLTAGE ....................................**

I hereby certify that the inspection/overhaul/repair/replacement/modification specified above has been carried out in accordance with the requirements of Chapter A4-3 of British Civil Airworthiness Requirements.

Signed ..................................................

Firm ......................................................

CAA Approval Ref.

Or Licence No. .......................................

Date ......................................................

# CHAPTER : 11
# POWER CONVERSION EQUIPMENT

In aircraft electrical installations a number of different types of consumer equipment are used which require power supplies different from those standard supplies provided by the main generator. For example, in an aircraft having a 28 volts d.c. primary power supply, certain instruments and electronic equipment are employed which require 26 volts and 115 volts a.c. supplies for their operation, and as we have already seen, d.c. cannot be entirely eliminated even in aircraft which are primarily a.c. in concept. Furthermore, we may also note that even within the items of consumer equipment themselves, certain sections of their circuits require different types of power supply and/or different levels of the same kind of supply. It, therefore, becomes necessary to employ not only equipment which will convert electrical power from one form to another, but also equipment which will convert one form of supply to a higher or lower value.

The equipment required for the conversion of main power supplies can be broadly divided into two main types, static and rotating, and the fundamentals of construction and operation of typical devices and machines are described blow

**STATICCONVERTINGEQUIPMENT**
The principal items under this category are : Rectifiers, transformers and static d.c. / a.c. converters.

**Rectifiers**
The process of converting an a.c. supply into a d.c. supply is known as rectification and any static apparatus used for this purpose is known as a rectifier.

The rectifying action is based on the principle that when a voltage is applied to certain combinations of metallic and nonmetallic elements in contact with each other, an exchange of electrons and positive current carriers (known as "holes") takes place at the contact surfaces. As a result of this exchange, a barrier layer is formed which exhibits different resistance and conductivity characteristics and allows current to flow through the element combination more easily in one direction than in the opposite direction. Thus, when the applied voltage is an alternating quantity the barrier layer converts the current into a unidirectional flow and provides a rectified output.

One of the elements used in combination is referred to as a "semiconductor" which by definition denotes that it possesses a resistivity which lies between that of a good conductor and a good insulator. Semiconductors are also further defined by the number of carriers, i.e. electrons and positive "holes", provided by the "crystal lattice" form of the element's atomic structure. Thus, an element having a majority of electron carriers is termed "n-type" while a semiconductor having a majority of "holes" is termed "p-type".

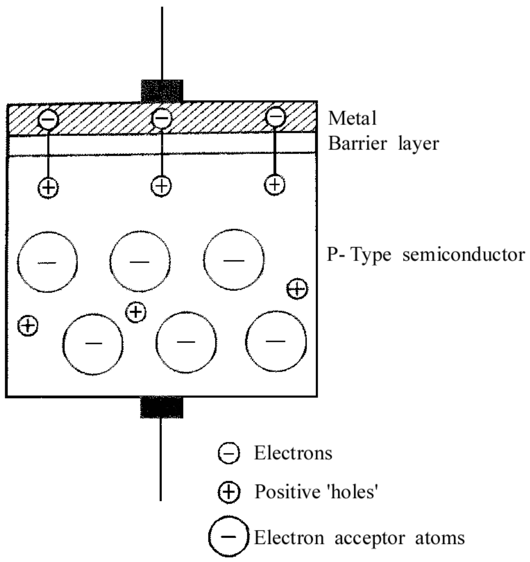



*Fig. 11.1, Semi-Conductor/metal junction.*

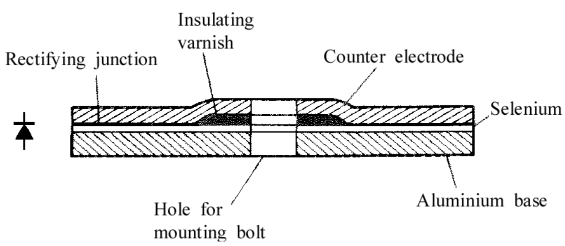



*Fig. 11.2, Cross-section of a selenium rectifier element.*

If a p-type semiconductor is in contact with a metal plate as shown in Fig. 11.1, electrons migrate from the metal to fill the positive holes in the semiconductor, and this process continues until the transfer of charge has established a p.d. sufficient to stop it. By this means a very thin layer of the semiconductor is cleared of positive holes and thus becomes an effective insulator, or barrier layer. When a voltage is applied such that the semiconductor is positive with respect to the metal, positive holes migrate from the body of the semiconductor into the barrier layer, thereby reducing its "forward" resistance and restoring conductivity. If, on the other hand, the semiconductor is made negative to the metal, further electrons are drawn from the metal to fill more positive holes and the "reverse" resistance of the barrier layer is thus increased. The greater the difference in the resistance to current flow in the two directions the better is the rectifying effect.

A similar rectifying effect is obtained when an n-type semiconductor is in contact with metal and a difference of potential is established between them, but in this case the direction of "easy" current flow is reversed. In practice, a small current does flow through a rectifier in the reverse direction because p-type material contains a small proportion of free electrons and n-type a small number of positive holes.

In the rectification of main a.c. power supplies, rectifiers are now invariably of the type employing the p-type nonmetallic semiconductors, selenium and silicon. Rectifiers employing germanium ( a metallic element) are also available but as their operating temperature is limited and protection against short duration overloads is difficult, they are not adopted in main power systems.

**Selenium Rectifiers**
The selenium rectifier is formed on an aluminium sheet which serves both as a base for the rectifying junction and as a surface for the dissipation of heat. A cross-section of an element is shown diagrammatically in Fig. 11.2 and from this it will be noted that the rectifying junction covers one side of the base with the exception of a narrow strip at the edges and a small area around the fixing hole which is sprayed with a layer of insulating varnish. A thin layer of a low-melting point alloy, referred to as the counter electrode, is sprayed over the selenium coating and insulating varnish. Contact with the two elements of the rectifying junction, or barrier layer, is made through the base on one side and the counter electrode on the other.

In practice a number of rectifying elements may be connected in series or parallel to form what is generally referred to as a rectifier stack. When connected in series the elements increase the voltage handling ability of a rectifier and when connected in parallel the ampere capacity is increased. (Fig. 11.3)

**Silicon Rectifiers**
Silicon rectifiers, or silicon junction diodes as they are commonly known, do not depend on such a large barrier layer as selenium rectifiers, and as a result they differ radically in both appearance and size. This is shown in figure 11.4.

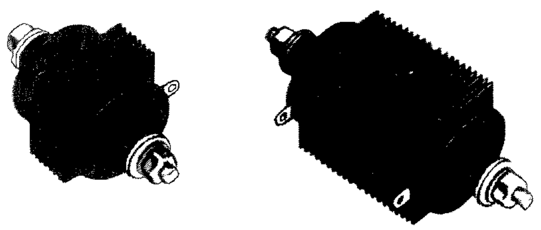


*Fig. 11.3, Typical rectifier stacks.*

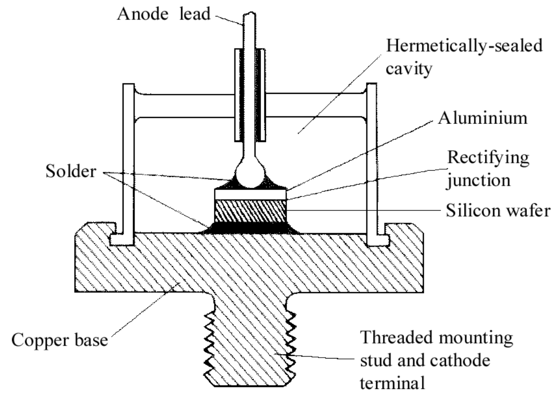



*Fig. 11.4, Silicon junction diode.*

This type of diode is used in constant frequency generator. The silicon is in the form of an extremely small slice cut from a single crystal and on one face it has a fused aluminium alloy contact to which is soldered an anode and lead. The other face is soldered to a base, usually copper, which forms the cathode and at the same time serves as a heat sink and dispatcher. The barrier layer is formed at the aluminium-silicon junction.

To protect the junction from water vapour and other deleterious materials, which can seriously impair its performance, it is mounted in a hermetically sealed case.

**OPERATING LIMITATIONS OF RECTIFIERS**
The limiting factors in the operation of a rectifier are (i) the maximum temperature permissible and (ii) the minimum voltage i.e. the reverse voltage, required to break down the barrier layer. In selenium rectifier the maximum temperature is of the order of 70° C, for germanium the temperature is about 50° C, while for silicon upto 150° may be reached without destroying the rectifier. It should be noted that these figure represent the actual temperature at the rectifying junction and, therefore, the rectifier, as a complete unit, must be at a much lower temperature.

**Silicon Controlled Rectifier (S. C. R.)**
An S. C. R. or thyristor is a development of a silicon diode and it has some of the characteristics of a thyratron tube. It is a three terminal device, two terminals corresponding to those of an ordinary silicon diode and the third, called the "gate" corresponding to the thyratron grid. The construction and operating characteristic is shown Fig. 11.5. The silicon wafer which is of " n type " has three more layers formed with in it in the sequence indicated.

When reverse voltage is applied an S. C. R. behaves in the same manner as a normal silicon diode, but when forward voltage is applied current flow is practically zero until a forward critical "break over" voltage is reached. The voltage at which break over takes place can be varied by applying small current signals between the gate and cathode, a method known as "firing". Once conduction has been initiated it can be stopped only by reducing the voltage to a very low value. The mean value of rectified voltage can be controlled, by adjusting the phasing of the gate signal with respect to the applied voltage. Thus an S.C.R. not only performs the function of power rectification, but also the function of an on off s/w, and a variable power output device. It is typical use is in battery charger unit.

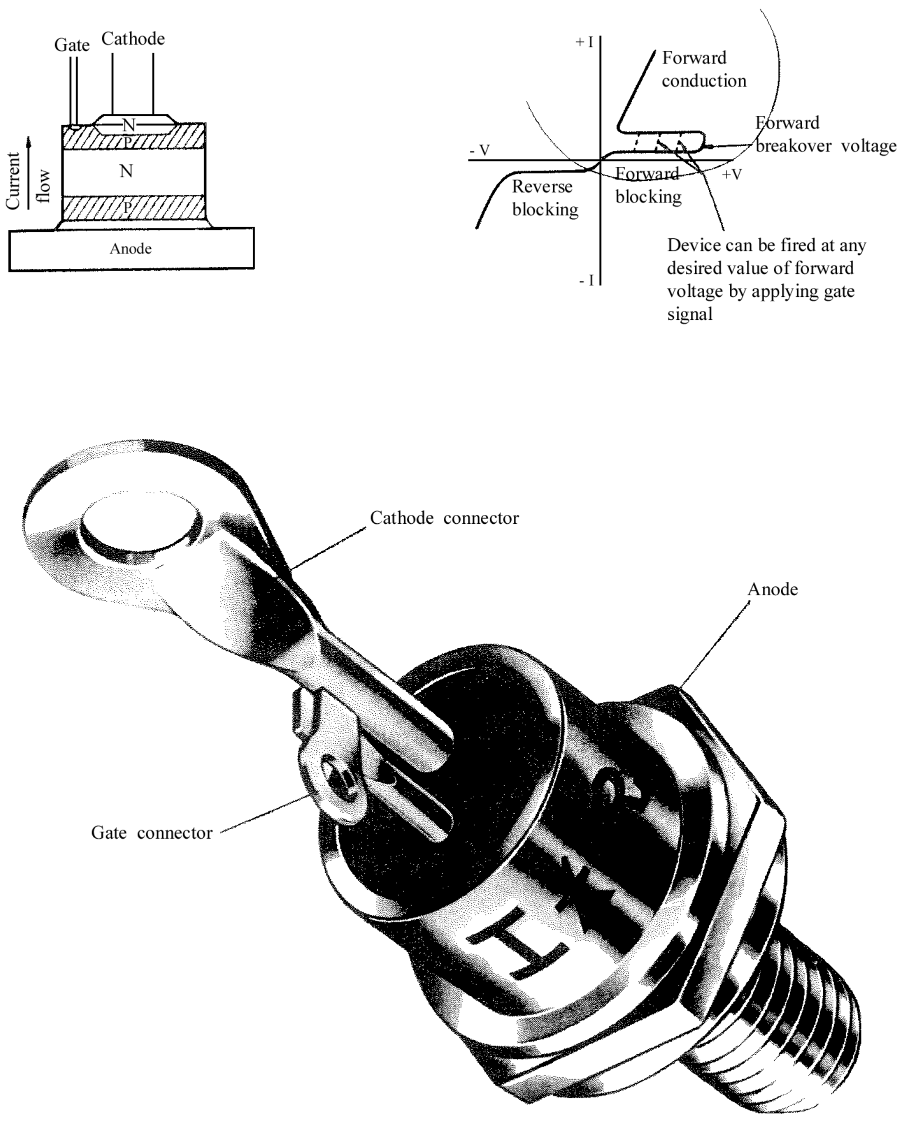



*Fig. 11.5, Silicon controlled rectifier.*

## TRANSFORMERS

A transformer is a device for converting a.c. at one frequency and voltage to a.c. at the same frequency but at another voltage. It consists of three main parts : (i) an iron core which provides a circuit of flow reluctance for an alternating magnetic field created by (ii) a primary winding which is connected to the main power source and (iii) a secondary winding which receives electrical energy by mutual induction from the primary winding and delivers it to the secondary circuit. There are two classes of transformers, voltage or power transformers and current transformers.

### Principle

The three main parts are shown in figure 11.6. when an alternative voltage is applied to the p.w. an a.c. will flow and by self inductance will establish a voltage in the s.w. which is opposite and almost equal to the applied voltage. The difference between these two voltages will allow just enough current (excitation current) to flow in the p.w to set up an alternating magnetic flux in the core. The flux cuts across the S.W. and by mutual inductance (in practice both windings are wound one on the other) a voltage is established in the S.W.

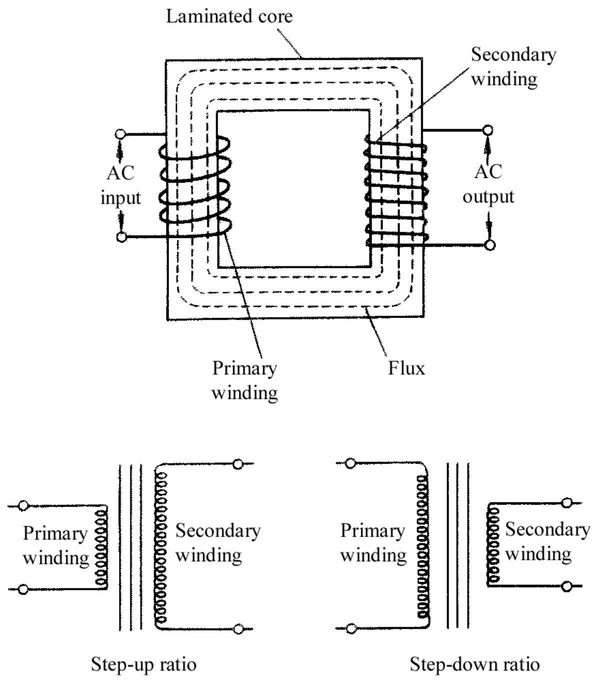



*Fig. 11.6, Transformer principle.*

When a load is connected to the S.W. terminals, the secondary voltage causes current to flow through the winding and a magnetic flux is produced which tends to neutralize the magnetic flux produced by the primary current. This in turn, reduces the self inductance, or opposition, voltage in primary winding, and allows more current to flow in it to restore the core flux to a value which is only very slightly less than the no-load value.

The primary current increases as the secondary load current increases, and decreases as the secondary load current decreases. When the load is disconnected, the primary winding current is again reduced to the small excitation current sufficient only to magnetize the core.

To accomplish the function of changing voltage from one value to another, one winding is wound with more turns than the other e.g., if the primary winding has 200 turns and the secondary 1000 turns, the voltage available at the secondary terminals will be 1000/200, or 5 times as great as the voltage applied to the primary winding. This ratio of turns ($N_2$) in the secondary to the number of turns ($N_1$) in the primary is coiled the turns or transformation ratio (r) and it is expressed by the equation.

$$r = N_2 / N_1 = E_2/E_1$$

where $E_1$ and $E_2$ are the respective voltages of the two windings.

When the transformation ratio is such that the transformer delivers a higher secondary voltage than the primary voltage it is said to be of the "step-up" type. Conversely, a "step-down" transformer is one which lowers the secondary voltage. The circuit arrangements for both types are also shown above.

**Construction Of Voltage Transformers**

The core of a voltage transformer is laminated and conventionally is built up of suitably shaped thin stamping, about 0.012 in. thick on average, of silicon-iron or nickel iron.

These materials have the characteristics of fairly high resistivity and low hysteresis, therefore, in the laminated form, the effects of both eddy currents and hysteresis are reduced to a minimum. Two different forms of constructions are in common use.

In one the laminations are L-shaped and are assembled to provide a single magnetic circuit, in this form it is used for the transformation of single phase a.c. The second known as the shell type, can be used for either single phase or three phase transformation and is one in which half the laminations are U-Shaped and the reminder are T shaped all of them being assembled to give a magnetic circuit with two paths. In both forms of construction the joints are staggered in order to minimize leakage at the joints. The laminations are held together by core clamps.

In some designs the cores are formed of strips are wound rather like a clock spring and bound together. The cores are then cut into two C-shaped parts to allow the prewound coils to be fitted. The mating surface of the two parts are often ground to give a very small effective gap which helps to minimize the excitation cement. After assembly of the windings the core parts are clamped together by a steel band around the outside of the core.

Transformer windings are of enamelled copper wire or strip and are normally wound on the core one upon the other, to obtain maximum mutual inductive effect, and are well insulated from each other. An exception to this normal arrangement is in a variant known as an auto transformer, in which the winding are in series and on a core made up of L shaped lamination, Parts of both primary and Secondary winding are wound on each side of the core. On the sheet type transformer both windings are wound on the centre limb for single phase operation and for three phase operation they are wound on each limb. Alternative tappings are generally provided on both windings of a transformer for different input and output voltage.

*Circuit Connection* Voltage transforms are connected so that the primary windings are in parallel with the supply voltage, the primary windings of current transformer as the name suggest is for the transformation of voltage from a single phase supply or form any one phase of three phase supply. Transformation of three phase a.c. can be carried out by means of three separate single phase transformers, or by a single three phase transformer. Transformers for three phase circuit can be connected in one of several combination of the star and delta connections, depending on the requirements for the transformer. The arrangements are shown above.

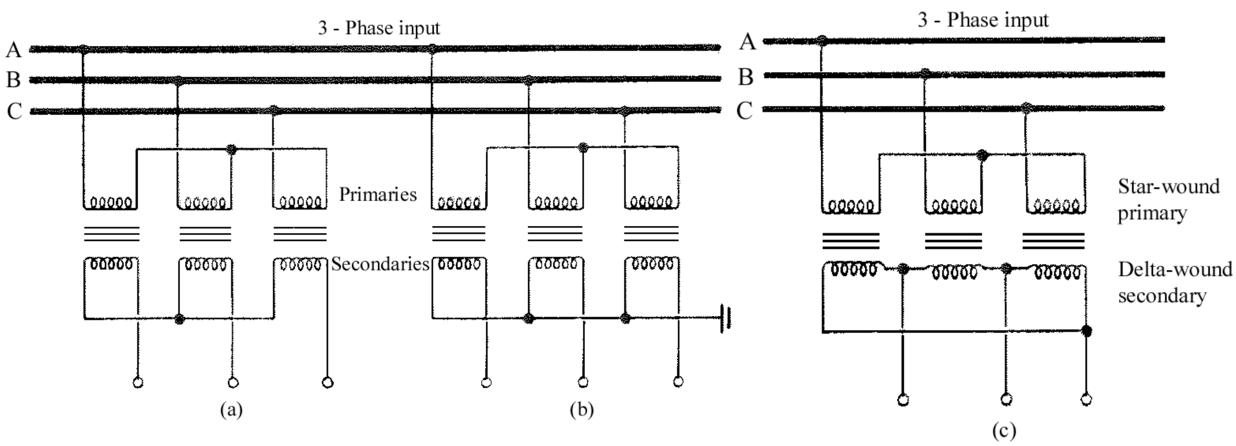



*Fig. 11.7, a. Star connection three wire.*
*b. Star connection four wire.*
*c. Star and delta connection.*

When the star connection is used in three phase transformers for the operation of three phase consumer equipment, the transformer may be connected as a three phase system fig 11. 7a. If a single phase loads have to be powered from a three phase supply, it is sometimes difficult to keep them balanced, it is therefore essential to provide a fourth or neutral wire so that connections of the loads may be made between this wire and any one of the three phase lines fig.11.7 b.

The interconnection of neutral points of two star windings is sometimes undesirable because this provides an external path for the flow of certain harmonic current which can lead to interference with radio communication equipment. This is normally overcome by connecting one of the two transformer windings delta, e.g., if the transformer supplies an unbalanced load, the primary winding is in star and sec is in delta as above (figure 11.7 c).

**Current Transformers**

Current transformers are used in many a.c. generator regulation and protection systems and also in conjunction with a.c. ammeters. These transformers have an input/output current relationship which is inversely proportional to the turns ratio of the primary and secondary windings. A typical unit is shown in fig.11.8. It is designed with only a SW on a toroidal strip wound core of silicon iron. The assembly together with the metal base is encapsulated in a resin compound moulding. The polarity of the transformer is indicated by the marking $H_1$ on the side facing the generator and $H_2$ on the side facing the load.

The primary winding is constituted by passing a main cable of the power system, through the core apex . The cable is wound with a single turn if it carries high currents, and with two or three turns if the carries low currents. The operating principle is the same as that of a conventional transformer.

Contrary to the practice adopted for voltage transformer, when ever the S.W. of current transformers are disconnected from their load circuits, terminals must be short circuited together. If this is not done, a dangerous voltage may develop which may be harmful to anyone accidentally touching the terminals, or may even cause an electrical breakdown between the windings.

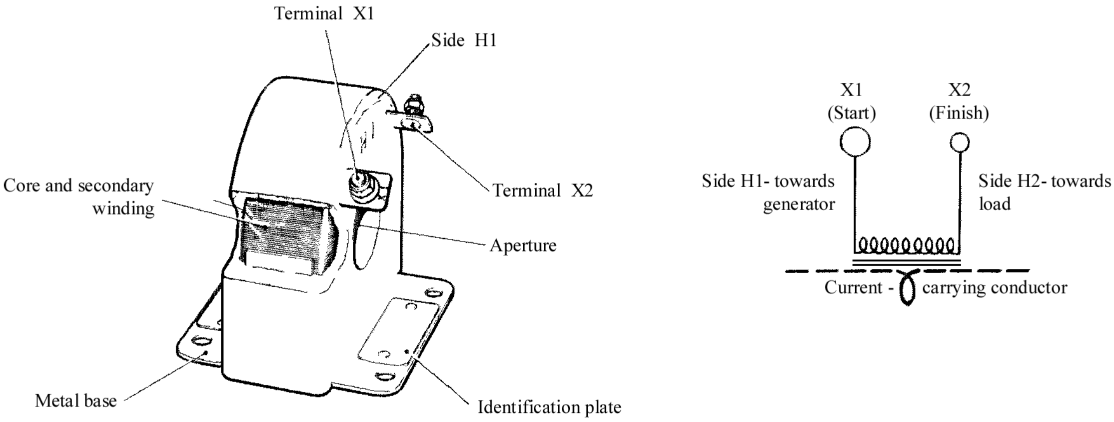



*Fig. 11.8, Current transformer.*

**Auto Transformers**

In circuit applications normally requiring only a small step up or step down of voltage, a special variant of transformer design is employed and this is known as an auto transformer. Circuit arrangement is shown 11.9, and from this it is seen that its most notable feature is that it consists of a single winding tapped to form primary and sec parts. In an example shown the tapping provide a sec output stepped up voltage output, since the number if primary turns is less than that of the sec turns.

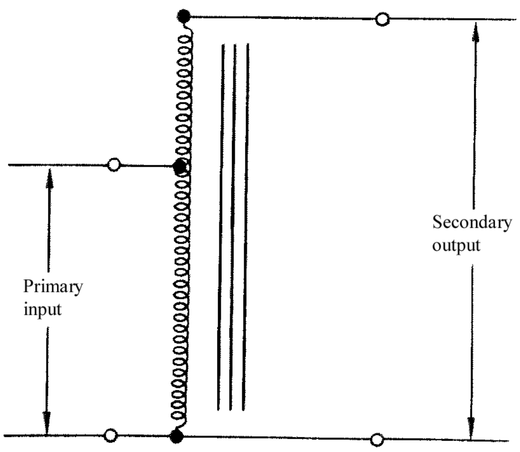



*Fig. 11.9, Circuit arrangement of an auto-transformer.*

When a voltage is applied to the primary terminals, current will flow through the position of winding spanned by these terminals. The magnetic flux due to this current will flow through the core and will, therefore, link with the whole of the winding. Those turns between the primary terminals act in the same way as the primary winding of a conventional transformer, and so they produce a self induction voltage in opposition to the applied voltage. The voltage induced in the remaining turns of the winding will be additive, there by giving a secondary output voltage greater than the applied voltage when a load circuit is connected to the S.T. a current due to the induced voltage will flow through the whole winding and will be in opposition to the primary current from the input terminals. Since the turns between the primary terminals are common to input and output circuits, alike they carry the difference between the induced current and primary current and they may, therefore, be wound with a smaller gauge wire than the remainder of the windings.

**Transformer Rectifier Unit**

Transformer rectifier units are combinations of static transformers and rectifiers, and are utilized in some a.c. systems as secondary supply units, and also as the main conversion units in aircraft having rectified a.c. power systems.

Transformer rectifier units are designed to operate on a regulated three phase input of 200 V at a frequency of 400 Hz and to provide a continuous of a output of 110 A at approximately 26V. The circuit is shown below.

The unit consists of a transformer and two three phase bridge rectifier assemblies mounted in separate sections of the casing. The transformer has a conventional star wound primary winding and secondary windings wound in star and delta. Each S.W. is connected to individual bridge rectifier assemblies made up of six silicon diodes and connected in parallel. An ammeter shunt (dropping 50mV at 100 A) is connected in the output side of the rectifier to enable current taken from the main d.c. output terminals to be measured at ammeter auxiliary terminals. These terminals, together with all others associated with input and output circuits, are grouped on a panel at one end of the unit cooling of the unit is by natural connection through gauge covered ventilation panels and in order to give warning of over heating conditions, thermal s/w 's are provided at the transformer and rectifier assemblies, and are connected to independent warning lights. The s/w 's are supplied with d.c. from an external source (normally one of the busbars) and their contacts close when temperature conditions at their respective locations rise to approx. 150°C and 200°C (Fig. 11.10 & Fig. 11.11).

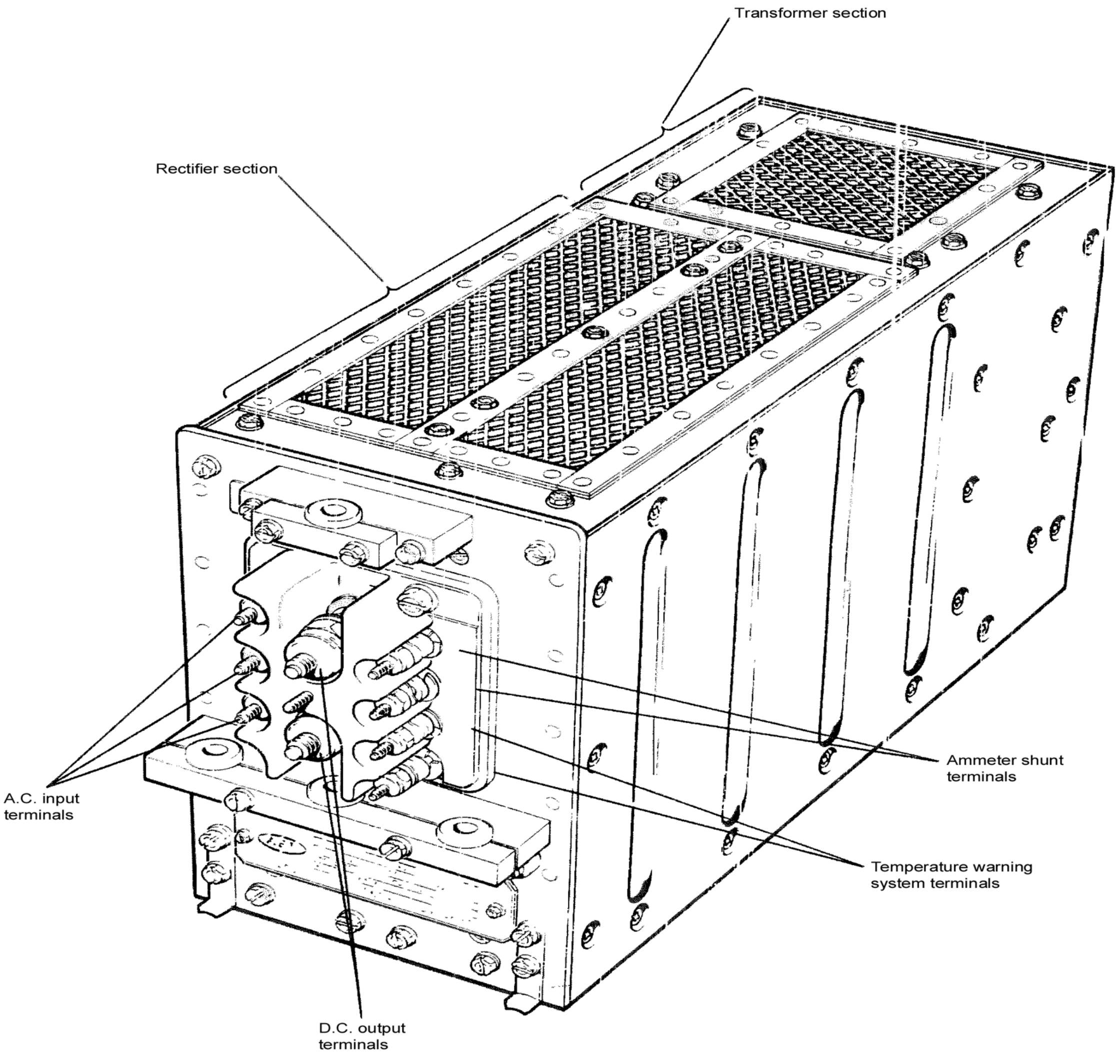



*Fig. 11.10, Transformer Rectifier Unit*

**Rotary Converting Equipment**

The most widely known device under this heading is the inverter designed to produce either 26 V or 115 V, 400 Hz a.c. depending on the secondary a.c. power requirements of an aircraft electrical system. Although now largely superseded by inverters of the solid state circuit or static type rotary inverters are still utilized in a number of smaller type aircraft.

A rotary inverter consists of a d.c. motor driving an a.c. generator, and since many of the systems which are to be operated from it are dependent on constant voltage and frequency the a.c. supply must be regulated accordingly. The methods of regulation may vary, but we may consider the commonly adopted method shown below in Fig. 11.12.

When the inverter is switched on, d.c. is supplied to the motor armature and shunt field winding, and also to the excitation field winding of the generator. Thus the motor starts driving the generator which will produce a three phase a.c. output at 115 volts. In order to control the voltage at this level, the d.c. supply is passed through a resistor in series with the generator field. This resistor is preset to give the required excitation current at the regulated d.c. system voltage level. Since the frequency of the generator output is dependent on speed then the present resistor is also connected in series with motor shunt field to provide sufficient excitation current to run the motor and generator at the speed necessary to provide at 400 Hg output.

Figure 11.13. illustrates a sectional view and circuit arrangement of another type of rotating inverter, and although it is only to be found on some older types of aircraft, it is an interesting example of variation in application of principles.

The motor and generator share a common immature and field system, and control of voltage and frequency is based on the carbon pile regulator principle.

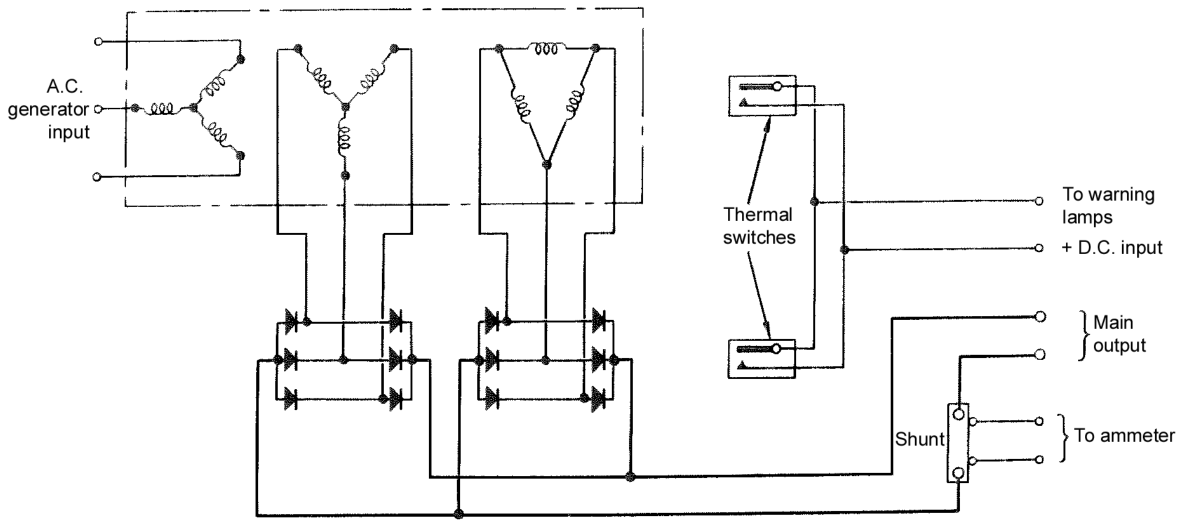



*Fig. 11.11, Schematic circuit of a transformer rectifier unit.*

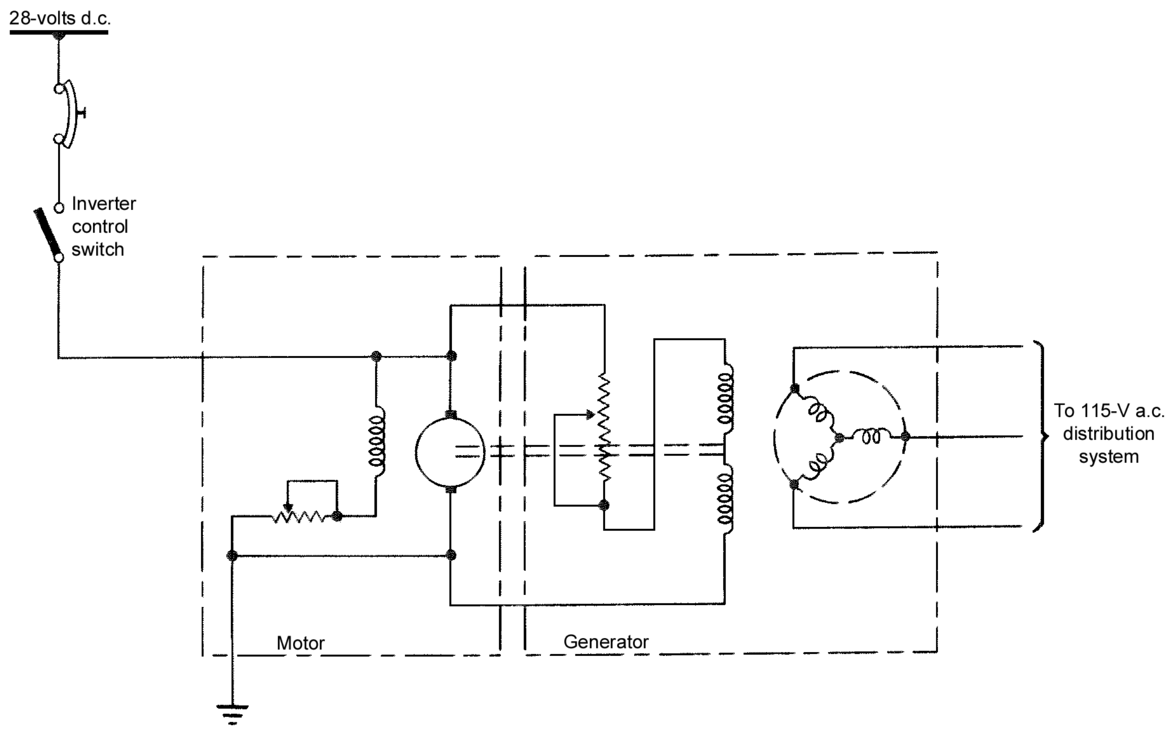



*Fig. 11.12, rotary inverter operation.*

The d.c. section of machine is of four pole compound wound type, the d.c. being supplied to the armature winding, series and shunt field windings. The a.c. section corresponds to a star a generator, the winding being located in the slots of the armature and beneath the d.c. winding. The a.c. winding is connected to a triple slip ring and bush gear assembly at the opposite end to the commutator. Thus, where the investor is in operation, a three phase output is induced in a rotating winding and not a fixed stator winding as in the case of a conventional a.c. generator.

The a.c. output is rectified and supplied to the voltage coil of the regulator which varies the pile resistance in the usual manner, this in turn, varying the current flow through the common field system to keep both the voltage and frequency of the a.c. output within limits.

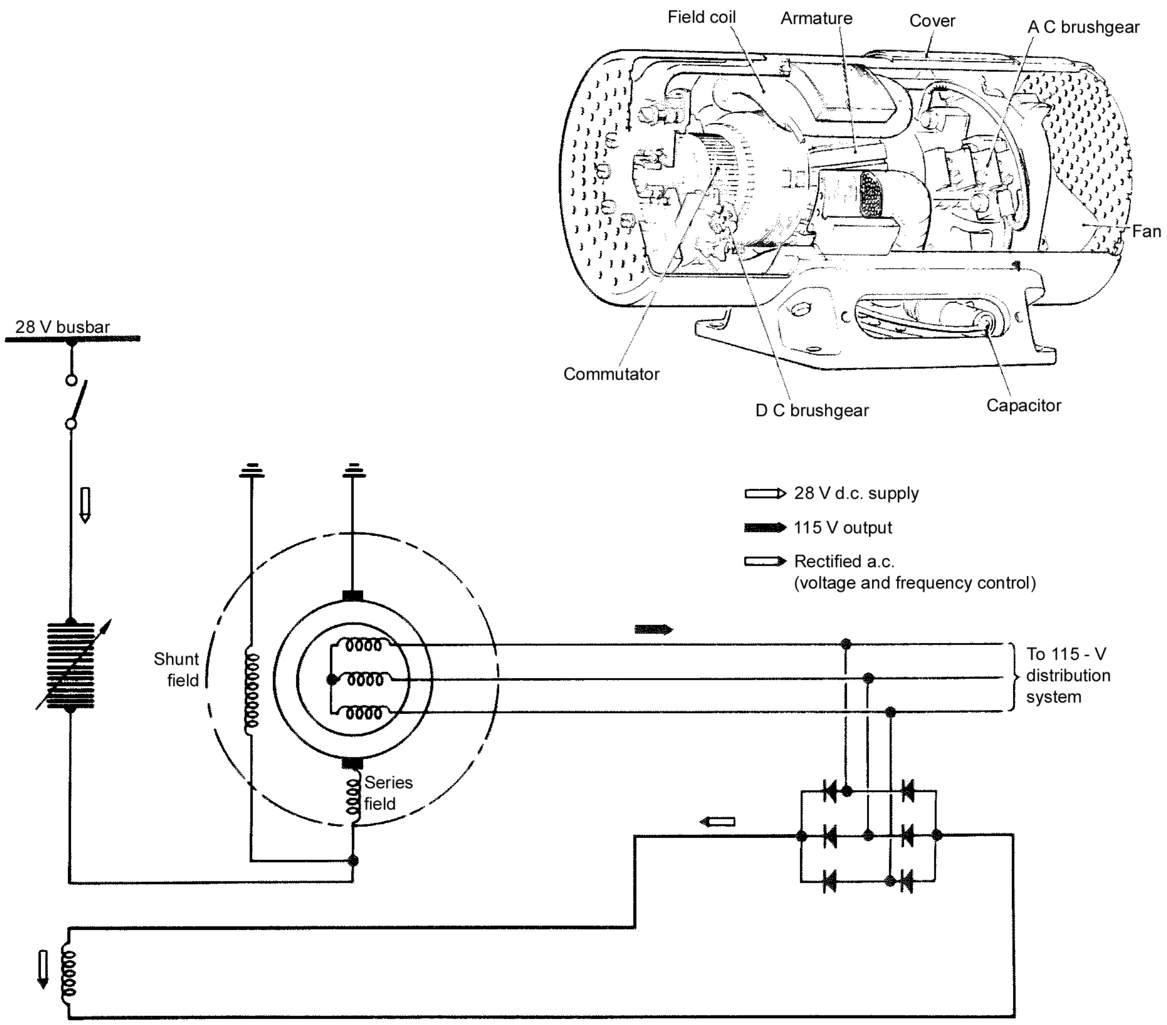



*Fig. 11.13, Rotary inverter (carbon pile regulator)*

**STATIC INVERTERS**

These inverters perform the same conversion function as the rotary machines, but by means of solid state or static circuit principles. They are employed in a number of types of aircraft in some cases as a normal sources of a.c. power but more usually to provide only emergency a.c. power to certain essential system when a failure of the normal 115 V source has occurred. The function of an inverter used for the conversion of battery supply to singles phase 115 V a.c. shown in block diagram, in fig.11.14.

The d.c. is supplied to transistorized circuits of a filter net work, a pulse shaper, a constant current generator power driver stage and output stage. After any variations in the input have been filtered or smoothed out, d.c. is supplied to a square wave generator which provides first stage conversion of the d.c. into square wave form a.c., and also establishes the required operating frequency of 400 Hz. This output is then supplied to a pulse shaper circuit which controls the pulse width of the signal and changes its wave form before it is passed on to the power driver stage. It will be noted from the diagram that the d.c. required for pulse shaper operation is supplied via a turn on delay circuit. The reason for this is to cause the pulse shaper to delay it output to the power driver stage until the voltage has stabilized. The power driver supplies a pulse width modulated symmetrical output to control the output stage the signal having a square wave from. The power driver also shorts itself out each time the voltage falls to zero i.e. during notch time.

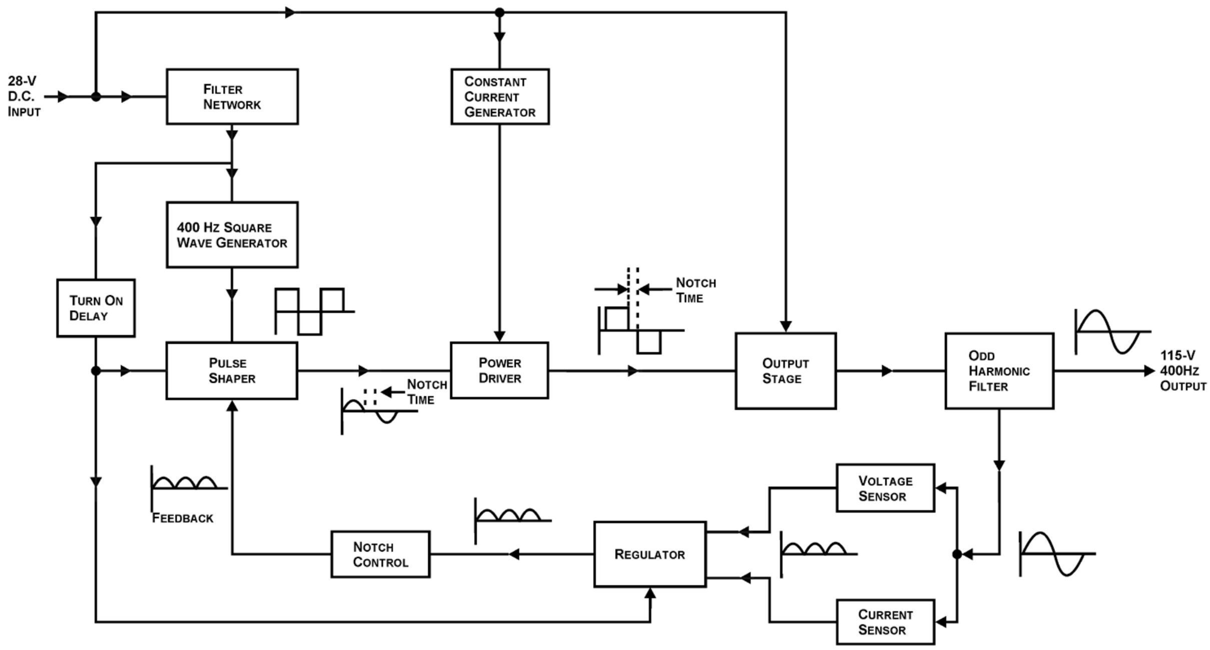



*Fig. 11.14, Static Inverter Principle.*

The output stage also produces a square wave output but of variable pulse width. This output is finally fed to a filter circuit which reduces the total odd harmonics to produce a sine wave output at the voltage and frequency required for operating systems connected to the inverters.

As in the case of other type of generators, the output of a static inverter must also be maintained within certain limits. In the example illustrated, this is done by means of a voltage sensor and a current sensor, both of which produce a rectified a.c. feed back signal which controls the 'notch time' of the pulse shaper output through the medium of a regulator circuit and a notch control circuit.

■■■

# CHAPTER : 12
# AIRCRAFT ELECTRICAL TEST EQUIPMENTS

## D.C. MEASURING INSTRUMENTS

Understanding the functional design and operation of electrical measuring instruments is very important, since they are used in repairing, maintaining, and troubleshooting electrical circuits. While some meters can be used for both d.c. and a.c. circuit measurement, only those used as d.c. instruments are discussed first. The meters used for a.c. or for both a.c. and d.c., are discussed latter on.

### Effects of Current

The effect of current may be classified as follows : chemical, physiological, photoelectric, piezoelectric, thermal, or electromagnetic.

### Chemical

When an electric current is passed through certain solutions, a chemical reaction takes place and a deposit forms on one electrode. The amount of this deposit is proportional to the amount of current. Industrially, this process is useful in electroplating and electrolysis. Although the chemical effect is useful in defining the standard ampere (the amount of current which causes .001118 grams of silver to be deposited in one second from a 15 per cent solution of silver nitrate), it is of no practical use in meters.

### Physiological

The physiological effect of current refers to the reaction of the human body to an electric current. An electric shock, although pannul at times, is too difficult to evaluate quantitatively and is, therefore, not practical for use in meters.

### Photoelectric

When electrons strike certain materials, a glow appears at the point of contact. The picture tube of a television set and the scope of a radar set illustrate this effect. Using the intensity of the light produced as a means of measuring the amount of current is neither accurate nor practical.

### Piezoelectric

Certain crystals such as quartz and Rochelle salts become deformed when a voltage is applied across two of the crystal faces. This effect is not visible to the human eye and is, therefore, impractical for use in meters.

### Thermal

When a current flows through a resistance, heat is produced. The amount of heat produced is equal to $I^2R$ . This relationship establishes that heat varies as the square of the current. Meters which employ the thermal effect in their operation are common.

### Electromagnetic

Whenever electrons flow through a conductor, a magnetic field proportional to the current is created. This effect is useful for measuring current and is employed in many practical meters.

The first four effects discussed are of no practical importance as electrical measuring devices. The last two effects, thermal and magnetic, are of practical use in metes. Since most of the meters in use have D'Arsonval movements, which operate because of the magnetic effect, only this type will be discussed in detail.

### D'Arsonval Meter

The basic d.c. meter movement is known as the D'Arsonval meter movement because it was first employed by the French scientist, D'Arsonval, in making electrical measurement. This type of meter movement is a current-measuring device which is used in the ammeter, voltmeter, and ohmmeter. Basically, both the ammeter and the voltmeter are current-measuring instruments, the principal difference being the method in which they are connected in a circuit. While an ohmmeter is also basically a current-measuring instrument, it differs from the ammeter and voltmeter in that it provides its own source of power and contains other auxiliary circuits.

### Ammeter

The D'Arsonval ammeter is an instrument designed for measuring direct current flowing in an electrical circuit and consists of the following parts : a permanent magnet, a moving element mounting, bearings, and a case which includes terminals, a dial, and screws. Each part and its function are described in the discussion which follows.

The permanent magnet furnishes a magnetic field which will react with the magnetic field set up by the moving element.

The moving element is mounted so that it is free to rotate when energized by the current to be measured. A pointer which moves across a calibrated scale is attached to this element. A moving coil mechanism is shown in figure 12.1. The controlling element is a spring, or springs, whose main function is to provide a counter or restoring force. The strength of this force increases with the turning of the moving element and brings the pointer to rest at some point on the scale. Two springs are generally used; they are wound in opposite directions to compensate for the expansion and contraction of the spring material due to temperature variation. The springs are made of nonmagnetic material and conduct current to and from the moving coil in some meters.

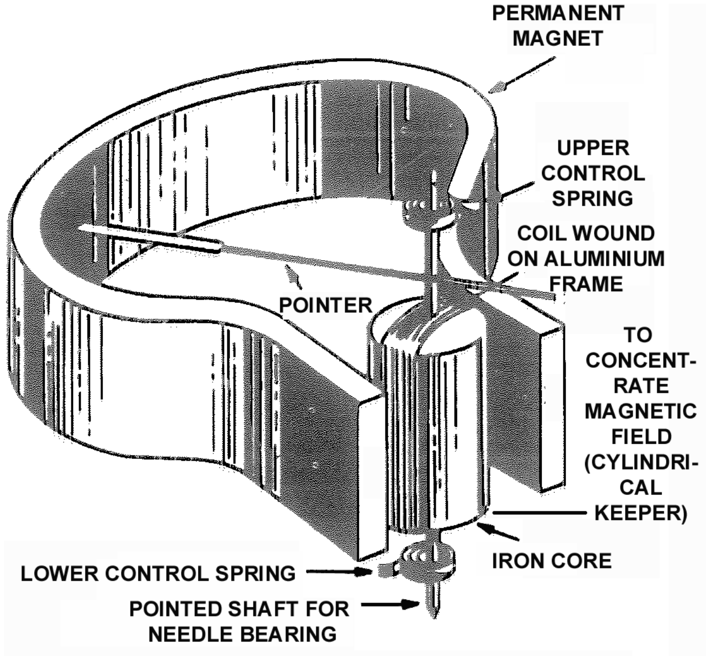



*Fig.12.1, Moving-coil element with pointer and springs.*

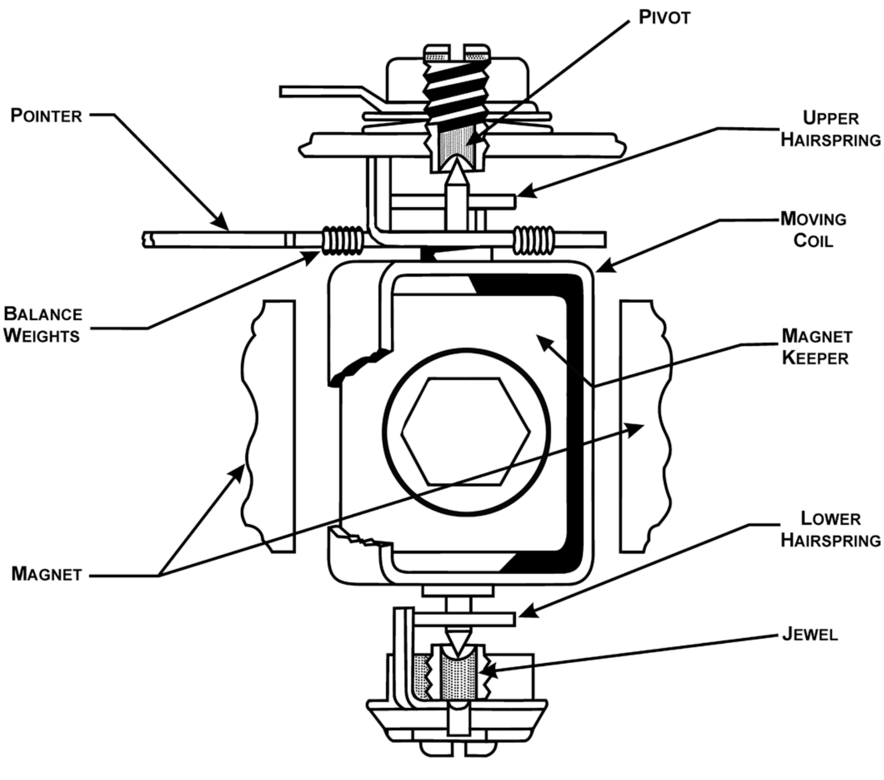



*Fig. 12.2, Method of mounting moving elements.*

The moving element consists of a shaft with very hard pivot points to carry the moving coil or other movable element (figure 12.1). The pivot points are so fitted into highly polished jewels or very hard glass bearings that the moving element can rotate with very little friction. Another type of mounting has been designed in which the pivot points are reversed and the bearings are inside the moving-coil assembly. A method of mounting moving elements is shown in figure 12.2.

The bearings are highly polished jewels such as sapphires, synthetic jewels, or very hard glass. These are usually round and have a conical depression in which the pivots rotate. They are set in threaded nuts which allow adjustment. The radius of the depression in the jewel is greater than the radius of the pivot point. This limits the area of contact surfaces and provides a bearing which, when operated dry, probably has the lowest constant friction value of any known type of bearing.

The case houses the instrument movement and protects it from mechanical injury and exposure. It also has a window for viewing the movement of the pointer across a calibrated scale. The dial has printed on it pertinent information such as the scale, units of measurement, and meter uses. The terminals are made of materials having very low electrical resistance. Their function is to conduct the required current into and away from the meter.

**Operation of the Meter Movement**

The major units are mounted in their relationship to one another (figure 12.3). Note that the coil portion of the moving element is in the magnetic field of the permanent magnet.

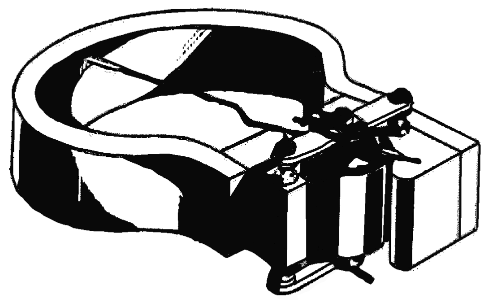


*Fig. 12.3, D'Arsonval meter movement.*

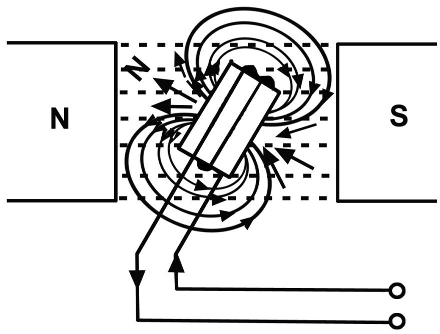



*Fig. 12.4, Effect of a coil in a magnetic field.*

In order to understand how the meter works, assume that the coil of the moving element is placed in a magnetic field as shown in figure 12.4.

The coil is pivoted so that it is able to rotate back and forth within the magnetic field set up by the magnet. When the coil is connected in a circuit, current flows through the coil in the direction indicated by the arrows and sets up a magnetic field within the coil. This field has the same polarity as the adjacent poles of the magnet. The interaction of the two fields causes the coil to rotate to a position so that the two magnetic fields are aligned. This force of rotation (torque) is proportional to the interaction between the like poles of the coil and the magnet and, therefore, to the amount of current flow in the coil. As a result, a pointer attached to the coil will indicate the amount of current flowing in the circuit as it moves across a graduated scale.

In the arrangement just discussed, note that any torque sufficient to overcome the inertia and friction of moving parts causes the coil to rotate until the fields align. This uncontrolled movement would cause inaccurate current readings. Therefore, the turning motion of the coil is opposed by two sprints. The value of the current flowing through the coil determines the turning force of the coil. When the turning force is equal to the opposition of the springs, the coil stops moving and the pointer indicates the current reading on a calibrated scale. In some meters the springs are made of conducting material and conduct current to and from the coil. The pole pieces of the magnet form a circular air tap within which the coil is pivoted.

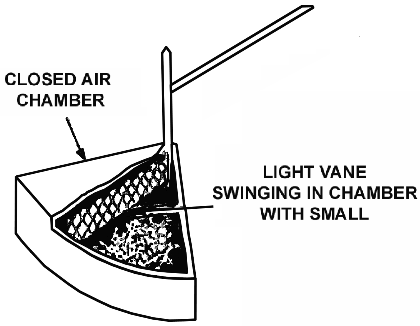



*Fig. 12.5, Air damping.*

To obtain a clockwise rotation, the north pole of the permanent magnet and that of the coil must be adjacent. The current flowing through the coil must, therefore, always be in the same direction. The D'Arsonval movement can be used only for d.c. measurements and the correct polarity must be observed. If the current is allowed to flow in the wrong direction through the coil, the oil will rotate counterclockwise and the pointer will be damaged. Since the movement of the coil is directly proportional to the current through the coil, the scale is normally a linear scale.

**Damping**

In order that meter readings can be made quickly and accurately, it is desirable that the moving pointer overshoot its proper position only a small amount and come to rest after not more than one or two small oscillations. The term "damping" is applied to methods used to bring the pointer of an electrical meter to rest after it has been set in motion. Damping may be accomplished by electrical means, by mechanical means, or by a combination of both.

**Electrical Damping**

A common method of damping by electrical means is to wind the moving coil on an aluminium frame. As the coil moves in the field of the permanent magnet, eddy currents are set up in the aluminium frame. The magnetic field produced by the eddy currents opposes the motion of the coil. The pointer will, therefore, swing more slowly to its proper position and come to rest quickly with very little oscillation.

**Mechanical Damping**

Air damping is a common method of damping by mechanical means. As shown in figure 12.5, a vane is attached to the shaft of the moving element and enclosed in an air chamber. The movement of the shaft is retarded because of the resistance which the air offers to the vane. Effective damping is achieved if the vane nearly touches the walls of the chamber.

**Meter Sensitivity**

The sensitivity of a meter movement is usually expressed as the amount of current required to give full-scale deflection. In addition, the sensitivity may be expressed as the number of millivolts across the meter when full-scale current flows through it. This voltage drop is obtained by multiplying the full-scale current by the resistance of the meter movement. A meter movement, whose resistance is 50 ohms and which requires 1 milliampere (mA.) for full-scale reading, may be described as a 50-millivolt 0-1 milliammeter.

**Extending the range of an Ammeter**

A 0-1 milliammeter movement may be used to measure currents greater than 1mA. by connecting a resistor in parallel with the movement. The parallel resistor is called a shunt because it by passes a portion of the current around the movement, extending the range of the ammeter. A schematic drawing of a meter movement with a shunt connected across it to extend its range is shown in figure 12.6.

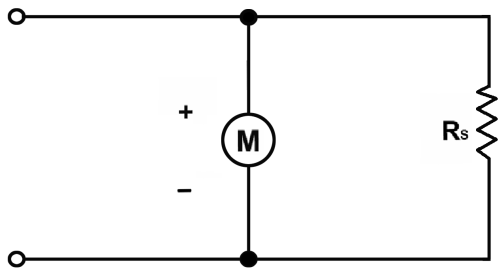



*Fig.12.6, Meter movement with shunt.*

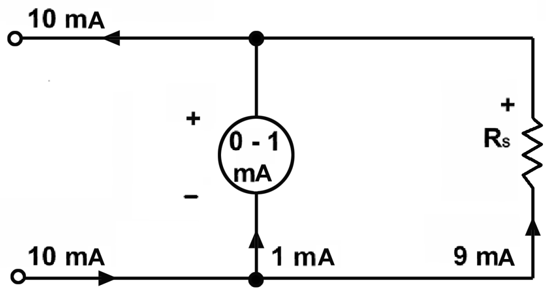



*Fig. 12.7, Circuit schematic for shunt resistor.*

**Determining the Value of a Shunt**

The value of a shunt resistor can be computed by applying the basic rules for parallel circuits. If a 50 millivolt 0-1 milliammeter is to be used to measure values of current up to 10 mA., the following procedure can be used : The first step involves drawing a schematic of the meter shunted by a resistor labelled $R_S$ (shunt resistor).

Since the sensitivity of the meter is known, the meter resistance can be computed. The circuit is then redrawn, and the branch currents can be computed, since a maximum of 1 mA. can flow through the meter. The voltage drop across $R_S$ is the same as that across the meter, $R_m$ :

$$E = IR_m$$
$$= 0.001 \times 50$$
$$= 0.050 \text{ volt.}$$

$R_s$ can be found by applying Ohm's law :

$$R_s = \frac{E_{rs}}{I_{rs}}$$

$$= \frac{0.050}{0.009}$$

$$= 5.55 \text{ ohms}$$

The value of the shunt resistor $(5.55\,\Omega)$ is very small, but this value is critical. Resistors used as shunts must have close tolerances, usually 1 per cent.

**Universal Ammeter Shunt**

The schematic drawing is figure 12.9, the universal shunt, shows an arrangement whereby two or more ranges are provided by tapping the shunt resistor at the proper points. In this arrangement, a 0-5 mA. movement with a resistance of 20 ohms is shunted to provide a 0-25 mA. range and a 0-50 mA . range.

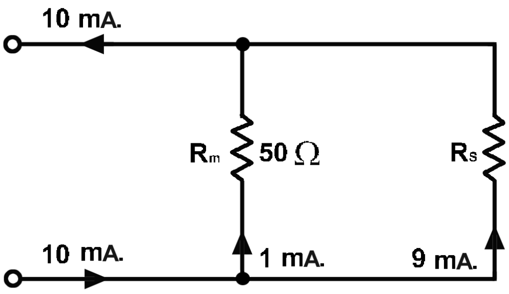



*Fig. 12.8, Equivalent meter circuit.*

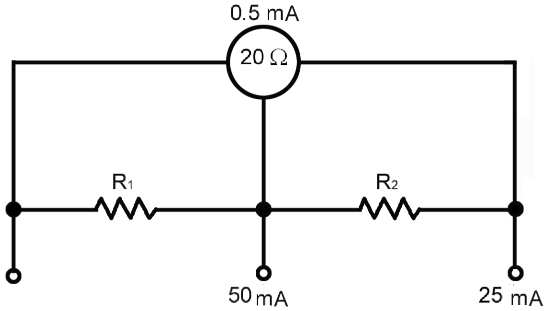



*Fig. 12.9, Universal ammeter shunt.*

Ammeters having a number of internal shunts are called multirange ammeters. A scale for each range is provided on the meter face (figure 12.10). Some multimeters avoid internal switching through the use of external shunts. Changing ammeter ranges involves the selection and installation on the meter case of the proper size shunt.

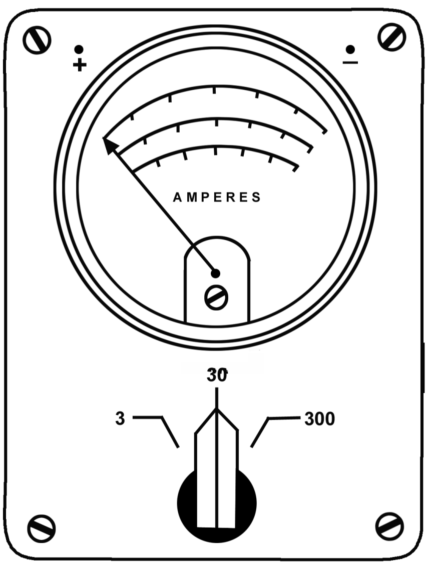



*Fig. 12.10, A multirange ammeter.*

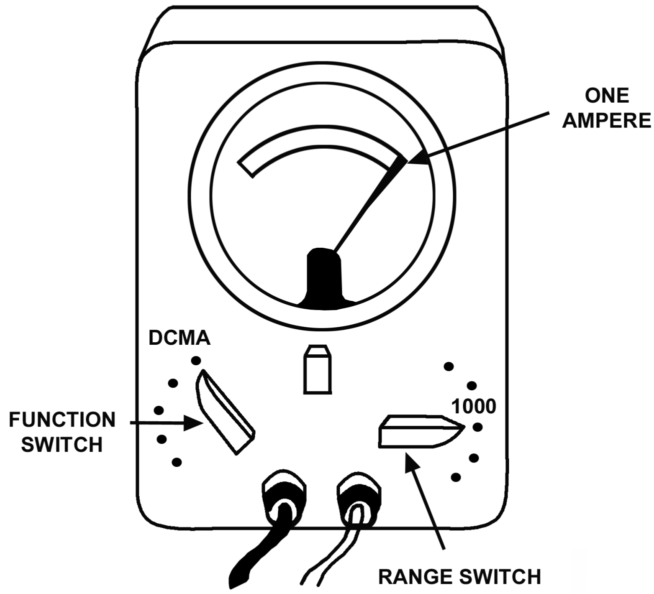



*Fig. 12.11, A multimeter set to measure one ampere.*

**MULTIMETERS**

Ammeters are commonly incorporated in multiple-purpose instruments such as multimeter or volt-ohm-milliammeters. These instruments vary somewhat according to the design used by different manufactures, but most incorporate the functions of an ammeter, a voltmeter, and an ohmmeter in one unit. A typical multimeter is shown in figure 12.11. This multimeter has two selector switches : a function switch and a range switch. Since a multimeter is actually three meters in one case, the function switch must be placed in proper position for the type of measurement to be made. In figure 12.11, the function switch is shown in the ammeter position to measure d.c. milliamperes and the range switch is set at 1000. Set in this manner, the ammeter can measure up to 1,000 milliamperes or 1 ampere.

Multimeters have several scales, and the one used should correspond properly to the position of the range switch. If current of unknown value is to be measured, always select the highest possible range to avoid damage to the meter. The test leads should always be connected to the meter in the manner prescribed by the manufacturer. Usually the red lead is positive and the black lead is negative, or common. Many multimeters employ color coded jacks as an aid in connecting the meter into the circuit to be tested. In figure 12.12, a multimeter properly set to measure current flow is connected into a circuit.

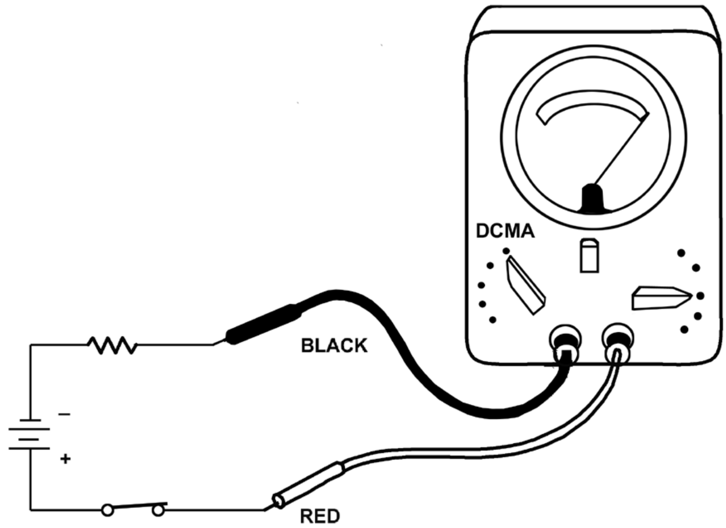



*Fig. 12.12, A multimeter set to measure current flow.*

The precautions to be observed when using an ammeter are summarized as follows :

1. Always connect an ammeter in series with the element through which the current flow is to be measured.
2. Never connect an ammeter across a source of voltage, such as a battery or generator. Remember that the resistance of an ammeter, particularly on the higher ranges, is extremely low and that any voltage, even a volt or so, can cause very high current or flow through the meter, causing damage to it.
3. Use a range large enough to keep the deflection less than full scale. Before measuring a current, form some idea of its magnitude. Then switch to a large enough scale or start with the highest range and work down until the appropriate scale is reached. The most accurate readings are obtained at approximately half-scale deflection. Many milliammeters have been ruined by attempts to measure amperes. Therefore, be sure to read the lettering either on the dial or on the switch positions and choose proper scale before connecting the meter in the circuit.
4. Observe proper polarity in connecting the meter in the circuit. Current must flow through the coil in a definite direction in order to move the indicator needle up-scale. Current reversal because of incorrect connection in the circuit results in a reversed meter deflection and frequently causes bending of the meter needle. Avoid improper meter connections by observing the polarity markings on the meter.

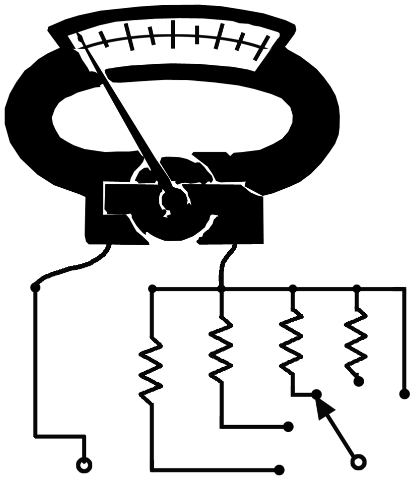


*Fig. 12.13, Simplified diagram of a voltmeter*

**VOLTMETER**

The D'Arsonval meter movement can be used either as an ammeter or a voltmeter (figure 12.13). Thus, an ammeter can be converted to a voltmeter by placing a resistance in series with the meter coil and measuring the current flowing through it. In other words, a voltmeter is a current-measuring instrument, designed to indicate voltage by measuring the current flow through a resistance of known value. Various voltage rages can be obtained by adding resistors in series with the meter coil. For low-range instruments, this resistance is mounted inside the case with the D'Arsonval movement and usually consists of resistance wire having a low temperature coefficient which is wound either on spools or card frames. For higher voltage ranges, the series resistance may be connected externally. When this is done, the unit containing the resistance is commonly called a multiplier.

**Extending the Voltmeter Range**

The value of the necessary series resistance is determined by the current required for full-scale deflection of the meter and by the range of voltage to be measured. Because the current through the meter circuit is directly proportional to the applied voltage, the meter scale can be calibrated directly in volts for a fixed series resistance.

For example, assume that the basic meter (microammeter) is to be made into a voltmeter with a full-scale reading of 1 volt. The coil resistance of the basic meter is 100 ohms, and 0.0001 ampere (100 microamperes) causes a full scale deflection. The total resistance, R, of the meter coil and series resistance is

$$R = \frac{E}{I} = \frac{1}{0.0001} = 10{,}000 \text{ ohms},$$

and the series resistance alone is

$$R_s = 10{,}000 - 100 = 9{,}900$$

Multirange voltmeters utilize one meter movement with the required resistances connected in series with the meter by a convenient switching arrangement. A multirange voltmeter circuit with three ranges is shown in figure 12.14. The total circuit resistance for each of the three ranges beginning with the 1-volt range is :

$$R = E\,I = \frac{1}{100} = 0.01 \text{ mega ohm}$$

$$\frac{100}{100} = 1 \text{ mega ohm}$$

$$\frac{1{,}000}{100} = 10 \text{ mega ohms}$$

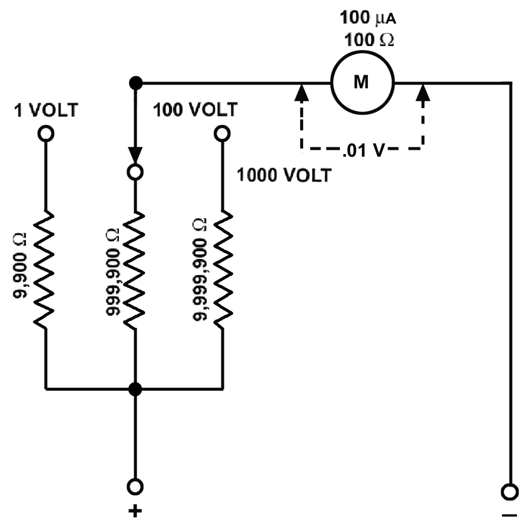



*Fig. 12.14, Multirange voltmeter schematic.*

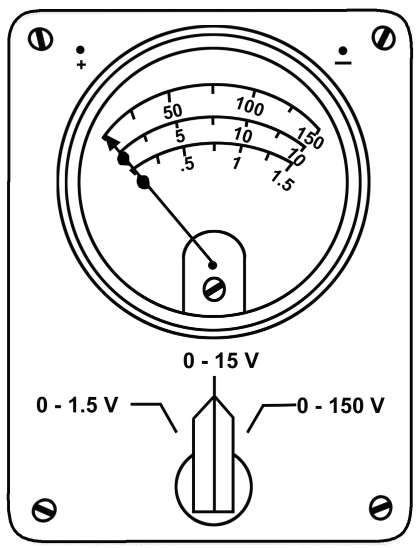



*Fig. 12.15, Typical multirange voltmeter.*

Multirange voltmeters, like multirange ammeters, are used frequently. They are physically very similar to ammeters, and their multipliers are usually located inside the meter with suitable switches or sets of terminals on the outside of the meter for selecting ranges (see figure 12.15).

Voltage-measuring instruments are connected across (in parallel with) a circuit. If the approximate value of the voltage to be measured is not known, it is best, as in using the ammeter, to start with the highest range of the voltmeter and progressively lower the range until a suitable reading is obtained.

In many cases, the voltmeter is not a central-zero indicating instrument. Thus, it is necessary to observe the proper polarity when connection the instrument to the circuit, as is the case when connecting the d.c. ammeter. The positive terminal of the voltmeter is always connected to the positive terminal of the source, and the negative terminal to the negative terminal of the source, when the source voltage is being measured. In any case, the voltmeter is connected so that electrons will flow into the negative terminal and out of the positive terminal of the meter. In figure 12.16 a multimeter is properly connected to a circuit to measure the voltage drop across a resistor. The function switch is set at the d.c. volts position and the range switch is placed in the 50-volt position.

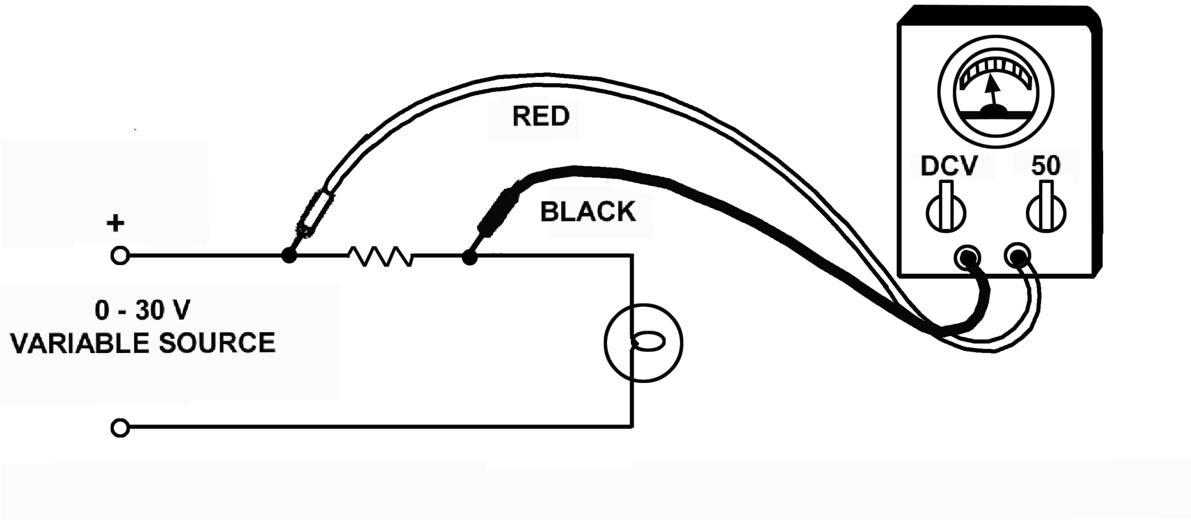



*Fig. 12.16, A multimeter connected to measure a circuit voltage drop.*

The function of a voltmeter is to indicate the potential difference between two point in a circuit. When the voltmeter is connected across a circuit, it shunts the circuit. If the voltmeter has low resistance, it will draw an appreciable amount of current. The effective resistance of the circuit will be lowered, and the voltage reading will consequently be lowered.

When voltage measurements are made in high resistance circuits, it is necessary to use a high-resistance voltmeter to prevent the shunting action of the meter. The effect is less noticeable in low-resistance circuits because the shunting effect is less.

**Voltmeter Sensitivity**

The sensitivity of a voltmeter is given in ohms per volt $(\Omega / E)$ and is determined by dividing the resistance $(R_m)$ of the meter plus the series resistance $(R_s)$ by the full-scale reading in volts.

Thus,

$$\text{Sensitivity} = \frac{R_m + R_s}{E}$$

This is the same as saying that the sensitivity is equal to the reciprocal of the current (in amperes); that is,

$$\text{Sensitivity} = \frac{\text{ohms}}{\text{volts}} = \frac{1}{\frac{\text{volts}}{\text{ohms}}} = \frac{1}{\text{amperes}}$$

Thus, the sensitivity of a 100-microampere movement is the reciprocal of 0.0001 ampere, or 10,000 ohms per volt.

The sensitivity of a voltmeter can be increased by increasing the strength of the permanent magnet, by using lighter weight materials for the moving element (consistent with increased number of turns on the coil), and by using sapphire jewel bearings to support the moving coil.

**Voltmeter Accuracy**

The accuracy of a meter is generally expressed in per cent. for example, a meter with an accuracy of 1 per cent will indicate a value within 1 per cent of the correct value. The statement means that, if the correct value is 100 units, the meter indication may be anywhere within the range of 99 to 101 units.

**OHMMETERS**

Two instruments are commonly used to check the continuity or to measure the resistance of a circuit or circuit element. These instruments are the ohmmeter and the megger, or megaohmmeter. The ohmmeter is widely used to measure resistance and to check the continuity of electrical circuits and devices. Its range usually extends to a few mega ohms. The megger is widely used for measuring insulation resistance, such as the resistance between the windings and frame of electric machinery, and for measuring the insulation resistance of cables, insulators, and bushings. Its range may extend to more than 1,000 mega ohms. When measuring very high resistances of this nature, it is not necessary to find the exact value or resistance, but rather to know that the insulation is either above or below a certain standard. When precision measurements are required, some type of bridge circuit is used. Ohmmeters may be of the series or shunt type.

**Series-type Ohmmeters**

A simplified schematic of an ohmmeter is shown in figure 12.17. E is a source of e.m.f.; $R_1$ is a variable resistor used to zero the meter; $R_2$ is a fixed resistor used to limit the current in the meter movement ; and A and B are test terminals across which the resistance to be measured is placed.

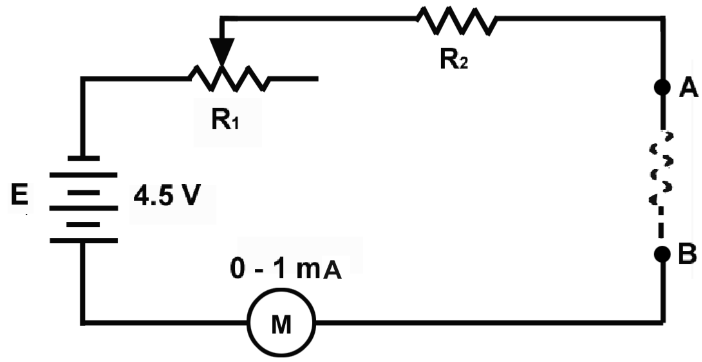



*Fig. 12.17, Ohmmeter circuit.*

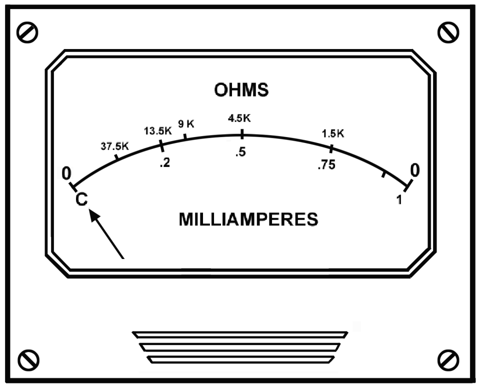



*Fig. 12.18, A typical ohmmeter scale.*

If A and B are connected together (short-circuited), the meter, the battery, and resistors $R_1$ and $R_2$ form a simple series circuit. With $R_1$ adjusted so that the total resistance in the circuit is 4,500 ohms, the current through the meter is 1 mA. and the needle deflects full scale. Since there is no resistance between A and B, this position of the needle is labelled zero (figure 12.18). If a resistance equal to 4,500 ohms is placed between terminals A and B, the total resistance is 9,000 ohms and the current is .5 mA.

This causes the needle to deflect half scale. This half-scale reading, labelled 4.5 K ohms, is equal to the internal resistance of the meter, in this instance 4,500 is placed between terminals A and B, the total resistance is 9,000 ohms and the current is .5 mA.

This causes the needle to deflect half scale. This half-scale reading, labelled 4.5 K ohms, is equal to the internal resistance of the meter, in this instance 4,500 ohms. If a resistance of 9,000 ohms is placed between terminals A and B, the needle deflects one-third scale. Resistance of 13.5 K and 1.5 K placed between terminals A and B will cause a deflection of one-fourth and three-fourths scale, respectively.

If terminals A and B are connected (open-circuited), no current flows and the needle does not move. The left side of the scale is, therefore, labelled infinity to indicate an infinite resistance.

A typical ohmmeter scale is shown in figure 12.18. Note that the scale is not linear and is crowded at the high resistance end. For this reason, it is good practice to use an ohmmeter range in which the readings are not too far from mid-scale. A good practice to use an ohmmeter range in which the readings are not too far from mid-scale. A good rule is to use a range in which the reading obtained does not exceed ten times, or is not less than one-tenth, the mid-scale reading. The useful range of the scale shown is, by this rule, from 450 ohms to 45,000 ohms.

Most ohmmeters have more than one scale Additional scales are made possible by using various values of limiting resistors and battery voltages. Some ohmmeters have a special scale called a low-ohm scale for reading low resistances. A shunt-type ohmmeter circuit is used for this scale.

**Shunt-Type Ohmmeter**

Shunt-type ohmmeters are used to measure small values of resistance. In the circuit shown in figure 12.19, E (voltage) is applied across a limiting resistor R and a meter movement in series. Resistance and battery values are chosen so that the meter movement deflects full scale when terminals A and B are open. When the terminals are short-circuited, the meter reads zero; the short circuit conducts all the current around the meter. The unknown resistance $R_x$ is placed between terminals A and B in parallel with the meter movement. The smaller the resistance value being measured, the less current flows through the meter movement.

The value of the limiting resistor R is usually made large compared to the resistance of the meter movement. This keeps the current drawn from the battery practically constant. Thus, the value of $R_x$ determines how much of this constant current flows through the meter and how much through $R_x$ .

Note that in a shunt-type ohmmeter, current is always flowing from the battery through the meter movement and the limiting resistor. Therefore, when using an ohmmeter with a low-ohm scale, do not leave the switch in low-ohm position.

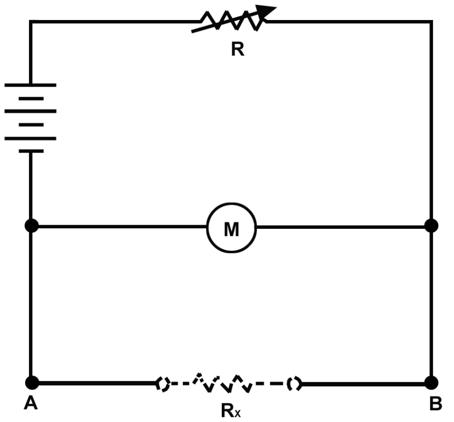



*Fig. 12.19, Shunt-type ohmmeter circuit.*

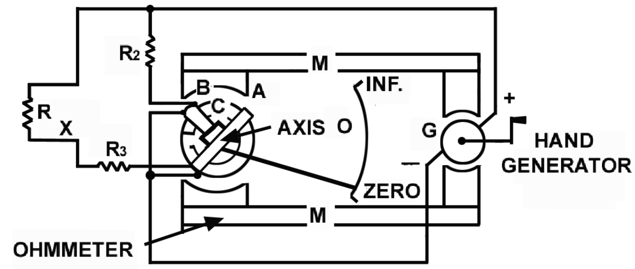



*Fig. 12.20, Simplified megger circuit.*

**Use of the Ohmmeter**

The ohmmeter is not as accurate a measuring device as the ammeter or the voltmeter because of the associated circuitry. Thus, resistance values cannot be read with greater than 5 to 10 per cent accuracy. While there are instruments which read the resistance of an element with very great accuracy, they usually are more complicated to use.

In addition to measuring the resistance, the ohmmeter is a very useful instrument for checking continuity in a circuit. Often, when troubleshooting electronic circuits or wiring a circuit, visual inspections of all parts of the current path cannot be readily accomplish. Therefore, it is not always apparent whether a circuit is complete or whether current might be flowing in the wrong part of the circuit because of contact with adjacent circuits. The best method of checking a circuit under these condition is to send a current through the circuit. The ohmmeter is the ideal instrument for checking circuits in this manner. It provides the power and the meter to indicate whether the current is flowing.

Observe the following precautions when using an ohmmeter :

1. Choose a scale which will contain the resistance of the element to be measured. In general, use a scale in which the reading will fall in the upper half of the scale (near full-scale deflection).

2. Short the lads together and set the meter to read zero ohms by setting the zero adjustment. If the scale is changed, readjust to zero ohms.

3. Connect the unknown resistance between the test leads and read its resistance from the scale. Never attempt to measure resistance in a circuit while it is connected to a source of voltage. Disconnect at least one end of the element being measured to avoid reading the resistance of parallel paths.

**Megger (Megaohmmeter)**

The megger, or megaohmmeter, is a high-range ohmmeter containing a hand-operated generator. It is used to measure insulation resistance and other high resistance values. It is also used for ground, continuity, and short-circuit testing of electrical power systems, The chief advantage of the megger over an ohmmeter is its capacity to measure resistance with a high potential, or "breakdown" voltage. This type of testing ensures that insulation or a dielectric material will not short or leak under potential electrical stress.

The megger (figure 12.20) consists of two primary elements, both of which are provided with individual magnetic fields from a common permanent magnet : (1) A hand-driven d.c. generator, G, which supplies the necessary current for making the measurement and (2) the instrument portion, which indicates the value of the resistance being measured. The instrument portion is of the opposed-coil type. Coils A and B are mounted on the movable member with a fixed angular relationship to each other and are free to turn as a unit in a magnetic field. Coil B tends to move the pointer counterclockwise and coil A, clockwise. The coils are mounted on a light, movable frame that is pivoted in jewel bearings and free to move about axis 0.

Coil A is connected in series with R3 and the unknown resistance, $R_x$ is connected between the + and - brushes of the d.c. generator. Coil B is connected in series with R2 and this combination is also connected across the generator. There are no restraining springs on the movable member of the instrument portion of the megger. When the generator is not in operation, the pointer floats freely and may come to rest at any position on the scale.

If the terminals are open-circuited, no current flows in coil A, and current in coil B alone controls the movement of the moving element. Coil B takes a position opposite the gap in the core (since the core cannot move and coil B can), and the pointer indicates infinity on the scale. When a resistance is connected between the terminals, current flows in coil A, tending to move the pointer clockwise. At the same time, coil B tends to move the pointer counterclockwise. Therefore, the moving element, composed of both coils and the pointer, comes to rest at a position at which the two forces are balanced. This position depends upon the value of the external resistance, which controls the relative magnitude of current of coil A. Because changes in voltage affect both coil A and B in the same proportion, the position of the moving element is independent of the voltage. If the terminals are short-circuited, the pointer rests at zero because the current in A is relatively large. The instrument is not damaged under these circumstances because the current is limited by R3. There are two types of hand-driven meggers : the variable type and the constant-pressure type. The speed of the variable-pressure megger is dependent on how fast the hand crank is turned. The constant-pressure megger utilizes a centrifugal governor, or slip clutch. The governor becomes effective only when the megger is operated at a speed above its slip speed, at which speed its voltage remains constant.

## A.C. MEASURING INSTRUMENTS

A d.c. meter, such as an ammeter, connected in an a.c. circuit will indicate zero, because the moving ammeter coil that carries the current to be measured is located in a permanent magnet field. Since the field of a permanent magnet remains constant and in the same direction at all times, the moving coil follows the polarity of the current. The coil attempts to move in one direction during half of the a.c. cycle and in the reverse direction during the other half when the current reverses.

The current reverses direction too rapidly for the coil to follow, causing the coil to assume an average position. Since the current is equal and opposite during each half of the a.c. cycle, the direct current meter indicates zero, which is the average value. Thus, a meter with a permanent magnet D'Arsonval meter may be used to measure alternating current or voltage if the current that passes through the meter is first rectified-that is, changed from alternating current to direct current.

### Rectifier A.C. Meters

Copper-oxide rectifiers are generally used with D'Arsonval d.c. meter movements to measure alternating currents and voltages; however, there are many types of rectifiers which may be used, some of which are included in the discussion of alternator systems.

A copper-oxide rectifier allows current to flow through a meter in only one direction. A shown in figure 12.21, the copper-oxide rectifier consists of copper-oxide disks separated alternately by copper disks and fastened together as a single unit. Current flows more readily from copper to copper oxide than from copper oxide to copper. When a.c. is applied, therefore, current flows in only one direction, yielding a pulsating d.c. output as shown by the output wave shapes in figure 12.22. This current can then be measured as it flows through the meter movement.

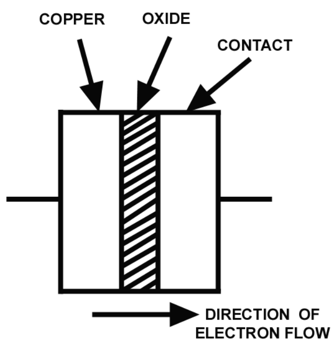



*Fig. 12.21, Copper-oxide rectifier.*

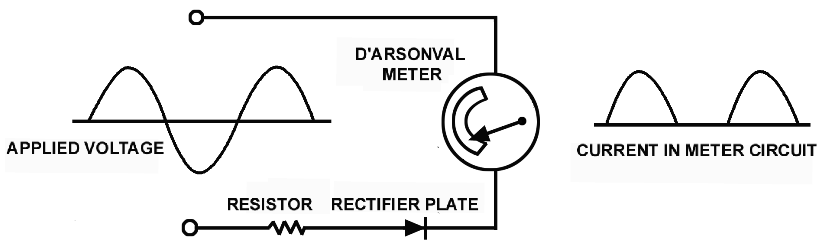



*Fig. 12.22, A half-wave rectifier circuit.*

In some a.c. meters, selenium or vacuum tube rectifiers are used in place of the copper-oxide rectifier. The principle of operation, however, is the same in all meters employing rectifiers.

### Electrodynamometer Meter Movement

The electrodynamometer meter can be used to measure alternating or direct voltage and current. It operates on the same principles as the permanent magnet moving-coil meter, except that the permanent magnet is replaced by an air-core electromagnet. The field of the electrodynamometer meter is developed by the same current that flows through the moving coils (see figure 12.23).

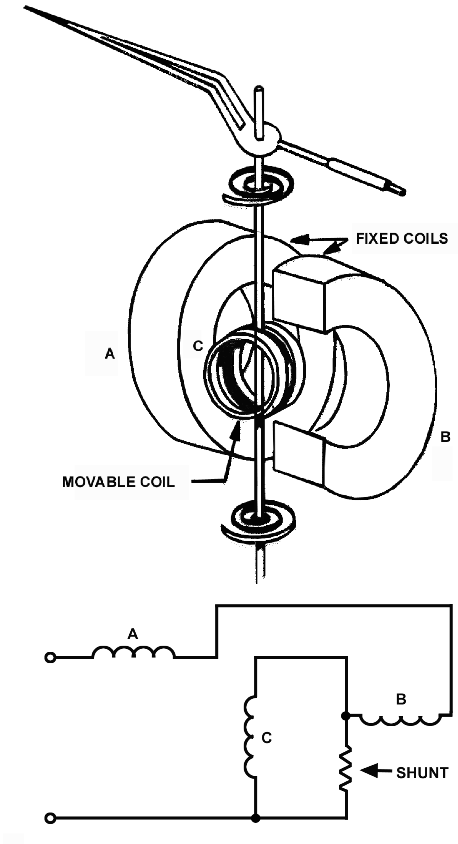



***Fig. 12.23, Simplified diagram of an electrodynamometer movement.***

In the electrodynamometer meter, two stationary field coils are connected in series with the movable coil. The movable coil is attached to the central shaft and rotates inside the two stationary field coils. The spiral springs provide the restraining force for the meter and also a means of introducing current to the movable coil.

When current flows through field coils A and B and movable coil C, coil C rotates in opposition to the springs and places itself parallel to the field coils. The more current flowing through the coils, the more the moving coil overcomes the opposition of the springs and the farther the pointer moves across the scale. If the scale is properly calibrated and the proper shunts or multipliers are used, the dynamometer movement will indicate current or voltage.

Although electrodynamometer meters are very accurate, they do not have the sensitivity of D'Arsonval meters and, for this reason, are not widely used outside the laboratory.

**Electrodynamometer Ammeter**

In the electrodynamometer ammeter, low resistance coils produce only a small voltage drop in the circuit measured. An inductive shunt is connected in series with the field coils. This shunt, similar to the resistor shunt used in d.c. ammeters, permits only part of the current being measured to flow through the coils. As in the d.c. ammeter, most of the current in the circuit flows through the shunt; but the scale is calibrated accordingly, and the meter reads the total current. An a.c. ammeter, like a d.c. ammeter, is connected in series with the circuit in which current is measured. Effective values are indicated by the meter. A schematic diagram of an electrodynamometer ammeter circuit is shown in figure 12.24.

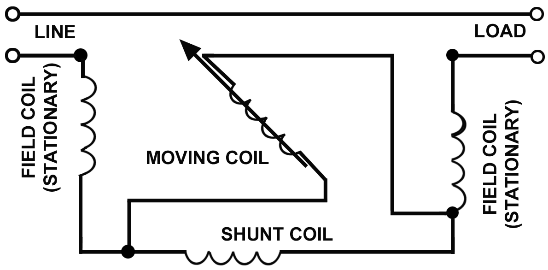



***Fig. 12.24, Electodynamometer ammeter circuit***

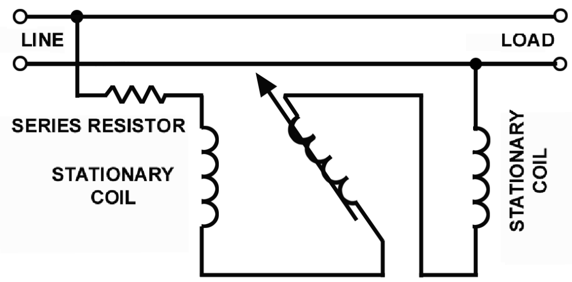



***Fig. 12.25, Electrodynamometer voltmeter circuit.***

**Electrodynamometer Voltmeter**

In the electrodynamometer voltmeter, field coils are wound with many turns of small wire. Approximately 0.01 ampere of current flow through both coils is required to operate the meter. Resistors of a noninductive material, connected in series with the coils, provide for different voltage ranges. Voltmeters are connected in parallel across the unit in which voltage is to be measured. The values of voltages indicated are effective values. A schematic diagram of an electrodynamometer voltmeter is shown in figure 12.25.

**Moving Iron-Vane Meter**

The moving iron-vane meter is another basic type of meter. It can be used to measure either a.c. or d.c. Unlike the D'Arsonval meter, which employs permanent magnets, it depends on induced magnetism for its operation. It utilizes the principle of repulsion between two concentric iron vanes, one fixed and one movable, places inside a solenoid, as shown in figure 12.26. A pointer is attached to the movable vane.

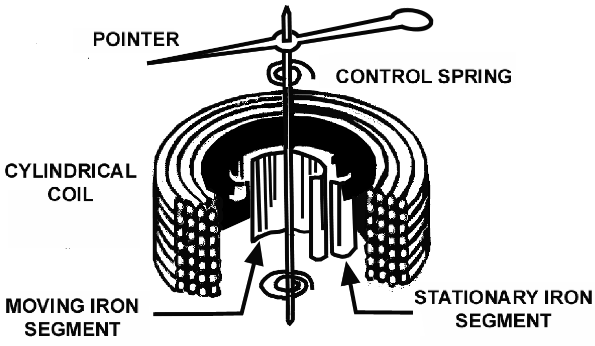



*Fig. 12.26, Moving iron-vane meter*

When current flows through the coil, the two iron vanes become magnetized with north poles at their upper ends and south poles at their lower ends for one direction of current through the coil. Because like poles repel, the unbalanced component of force, tangent to the movable element, causes it to turn against the force exerted by the springs.

The movable vane is rectangular in shape and the fixed vane is tapered. This design permits the use of relatively uniform scale.

When no current flows through the coil, the movable vane is positioned so that it is opposite the larger portion of the tapered fixed vane, and the scale reading is zero. The amount of magnetization of the vanes depends on the strength of the field, which, in turn, depends on the amount of current flowing through the coil. The force of repulsion is greater opposite the larger end of the fixed vane than it is nearer the smaller end. Therefore, the movable vane moves toward the smaller end through an angle that is proportional to the magnitude of the coil current. The movement ceases when the force of repulsion is balanced by the restraining force of the spring.

Because the repulsion is always in the same direction (toward the smaller end of the fixed vane), regardless of the direction of current flow through the coil, the moving iron-vane instrument operates on either d.c. or a.c. circuits.

Mechanical damping in this type of instrument can be obtained by the use of an aluminium vane attached to the shaft so that, as the shaft moves, the vane moves in a restricted air space.

When the moving iron-vane meter is designed to be used as a voltmeter, the solenoid is wound with many turns of small wire. Portable voltmeters are made with self-contained series resistance for ranges up to 750 volts. Higher ranges are obtained by the use of additional external multipliers.

The moving iron-vane instrument may be used to measure direct current but has an error due to residual magnetism in the vanes. The error may be minimized by reversing the meter connections and averaging the readings. When used on a.c. circuits the instruments has an accuracy of 0.5 per cent. Because of its simplicity, its relatively low cost, and the fact that no current is conducted to the moving element, this type of movement is used extensively to measure current and voltage in a.c. power circuits. However, because the reluctance of the magnetic circuit is high, the moving iron-vane meter requires much more power to produce full-scale deflection than is required by a D'Arsonval meter of the same range. Therefore, the moving iron-vane meter is seldom used in high-resistance low-power circuits.

**Inclined-Coil Iron-Vane Meter**

The principle of the moving iron-vane mechanism is applied to the inclined-coil type of meter, which can be used to measure both a.c. and d.c. The inclined-coil, iron vane meter has a coil mounted at angle to the shaft. Attached obliquely to the shaft, and located inside the coil, are two soft-iron vanes. When no current flows through the coil, a control spring holds the pointer at zero, and the iron vanes lie in planes parallel to the plane of the coil. When current flows through the coil, a control spring holds the pointer at zero, and the iron vanes lie in planes parallel to the plane of the coil. When current flows through the coil, the vanes tend to line up with magnetic lines passing through the center of the coil at right angles to the plane of the coil. Thus, the vanes rotate against the spring action to move the pointer over the scale.

The iron vanes tend to line up with the magnetic lines regardless of the direction of current flow through the coil. Therefore, the inclined-coil, iron-vane meter can be used to measure either alternating current or direct current. The aluminium disk and the drag magnets provide electromagnetic damping.

Like the moving iron-vane meter, the inclined coil type requires a relatively large amount of current for full-scale deflection and is seldom used in high-resistance low-power circuits.

As in the moving iron-vane instruments, the inclined-coil instrument is wound with few turns of relatively large wire when used as an ammeter and with many turns of small wire when used as a voltmeter.

**Thermocouple meter**

If the ends of two dissimilar metals are welded together and this junction is heated, a d.c. voltage is developed across the two open ends. The voltage developed depends on the material of which the wires are made and on the difference in temperature between the heated junction and the open ends.

In one type of instrument, the junction is heated electrically by the flow of current through a heater element. It does not matter whether the current is alternating or direct because the heating effect is independent of current direction. The maximum current that can be measured depends on the current rating of the heater, the heat that the thermocouple can stand without being damaged, and on the current rating of the meter used with the thermocouple. Voltage can also be measured if a suitable resistor is placed in series with the heater. In meter applications, a D'Arsonval meter is used with a resistance wire heater, as shown in figure 12.27.

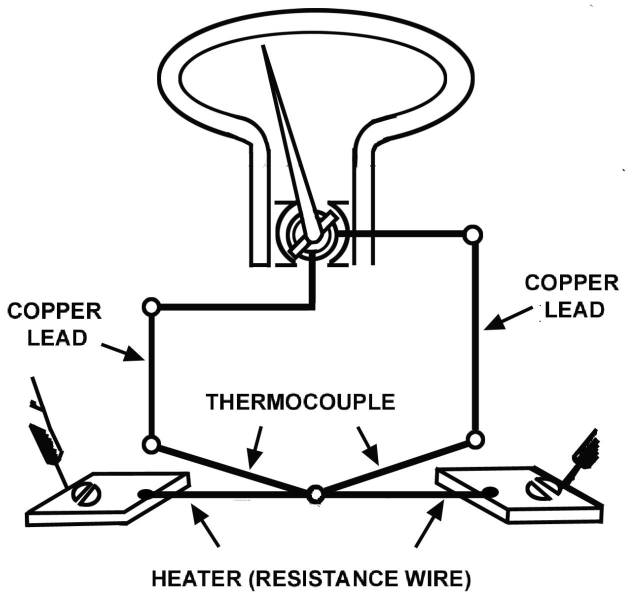



*Fig. 12.27, Simplified diagram of a thermocouple meter.*

As current flows through the resistance wire, the heat developed is transferred to the contact point and develops an e.m.f. which causes current to flow through the meter. The coil rotates and causes the pointer to move over a calibrated scale. The amount of coil movement is dependent on the amount of heat, which varies as the square of the current. Thermocouple meters are used extensively in a.c. measurements.

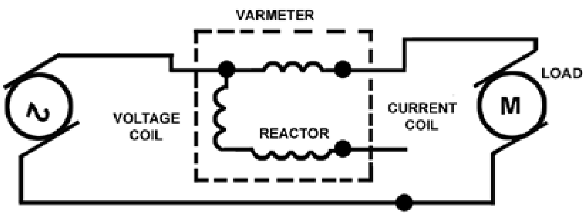



*Fig. 12.28, A varmeter connected in an a.c. circuit.*

**Varmeters**

Multiplying the volts by the amperes in an a.c. circuit gives the apparent power : the combination of the true power which does the work and the reactive power which does no work and is returned to the line. Reactive power is measured in units of vars (volt-amperes reactive) or kilovars (kilo-volt-amperes reactive, abbreviated KVAR). When properly connected, wattmeters measure the reactive power. As such, they are called varmeters. The illustration in figure 12.28 shows a varmeter connected in an a.c. circuit.

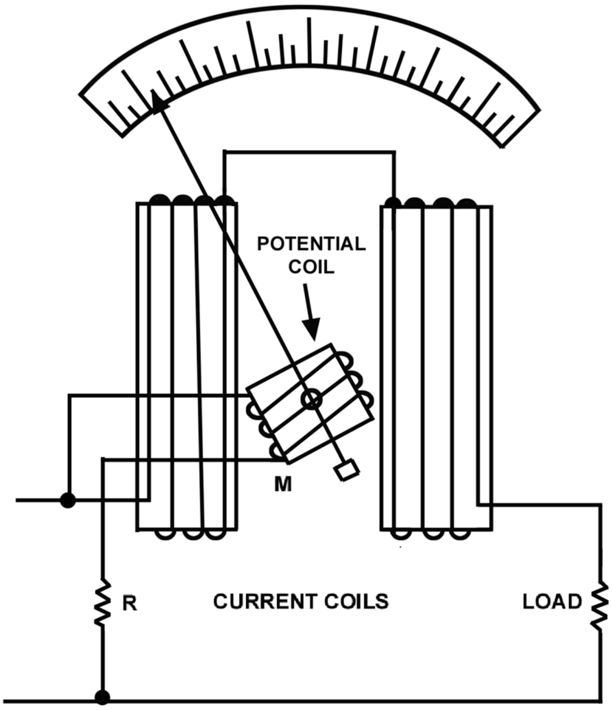



*Fig. 12.29, Simplified electrodynamometer, wattmeter circuit.*

## Wattmeter

Electric power is measured by means of a wattmeter. Because electric power is the product of current and voltage, a wattmeter must have two elements, one for current and the other for voltage, as indicated in figure 12.29. For this reason, wattmeters are usually of the electrodynamometer type.

The movable coil with a series resistance forms the voltage element, and the stationary coils constitute the current element. The strength of the field around the potential coil depends on the amount of current that flows through it. The current, in turn, depends on the load voltage applied across the coil and the high resistance in series with it. The strength of the field around the current coils depends on the amount of current flowing through the lad. Thus, the meter deflection is proportional to the product of the voltage across the potential coil and the current through the current coils. The effect is almost the same (if the scale is properly calibrated) as if the voltage applied across the load and the current through the load were multiplied together.

If the current in the line is reversed, the direction of current in both coils and the potential coil is reversed, the net result is that the pointer continues to read up-scale. Therefore, this type of wattmeter can be used to measure either a.c. or d.c. power.

## FREQUENCY METERS

Alternating-current electrical equipment is designed to operate within a given frequency range. In some instances the equipment is designed to operate at one particular frequency, as are electric clocks and time switches. For example, electric clocks are commonly designed to operate at 60 c.p.s. If the supply frequency is reduced to 59 c.p.s, the clock will lose one minute every hour.

Transformers and a.c. machinery are designed to operate at a specified frequency. If the supply frequency falls more than 10 per cent from the rated value, the equipment may draw excessive current, and dangerous overheating will result. It is, therefore, necessary to control the frequency of electric power systems. Frequency meters are employed to indicate the frequency so that corrective measures can be taken if the frequency varies beyond the prescribed limits.

Frequency meters are designed so that they will not be affected by changes in voltage. Because a.c. systems are designed to operate normally at one particular frequency, the range of the frequency meter may be restricted to a few cycles on either side of the normal frequency. There are several types of frequency meters, including the vibrating-reed type. Of these types, the vibrating-reed frequency meter is used most often in aircraft systems, and is discussed in some detail.

**Vibrating-Reed Frequency Meter**

The vibrating-reed type of frequency meter is one of the simplest devices for indicating the frequency of an a.c. source. A simplified diagram of one type of vibrating-reed frequency meter is shown in figure 12.30.

The current whose frequency is to be measured flows through the coil and exerts maximum attraction on the soft-iron armature twice during each cycle (A of figure 12.30). The armature is attached to the bar, which is mounted on a flexible support. Reeds of suitable dimensions to have natural vibration frequencies of 110, 112, 114, and so forth up to 130 c.p.s. are mounted on the (B of figure 5.30). The reed having a frequency of 110 c.p.s. is marked "55" cycles; the one having a frequency of 130 c.p.s. is marked "65" c.p.s., and so forth.

In some instruments the reeds are the same lengths, but are weighted by different amounts at the top so that they will have different natural rates of vibration.

When the coil is energized with a current having a frequency between 55 and 65 c.p.s., all the reeds are vibrated slightly; but the reed having a natural frequency closest to that of the energizing current (whose frequency is to be measured) vibrates through a larger amplitude. The frequency is read from the scale value opposite the reed having the greatest amplitude of vibration.

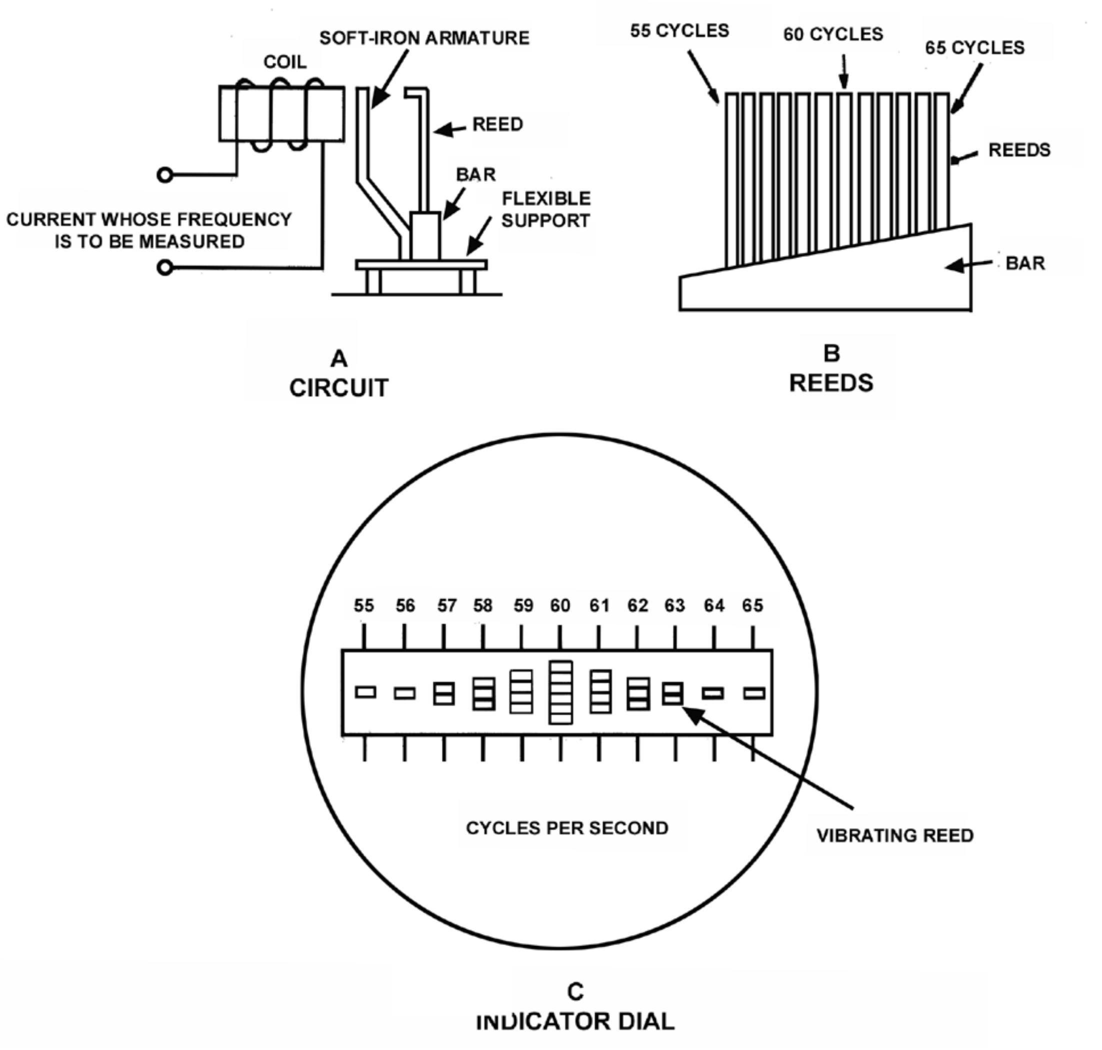



*Fig. 12.30, Simplified diagram of a vibrating-reed frequency meter.*

An end view of the reeds is shown in the indicator dial (C of figure 5.30). If the energizing current has a frequency of 60 c.p.s., the reed marked "60" c.p.s. will vibrate the greatest amount, as shown.

**INTRODUCTION**

This chapter gives general guidance on the inspection and testing of bonding and electrical circuits after installation and at the periods specified in the approved Maintenance Schedule for the aircraft concerned.

The methods of testing and inspection will vary with different types of aircraft and the equipment fitted, therefore reference must be made to the appropriate Maintenance Manuals for detailed information.

**GENERAL**

Each test requires specified equipment and care should be taken that it is correctly used (e.g. good electrical contact should always be made).

To ensure the reliability of test equipment, it should be carefully serviced and certified at the periods recommended by the manufacturer. The performance of equipment should also be checked before and after use.

After completion of all tests, the installations should be inspected to ensure that all connections have been remade and secured, and that test equipment, tools, etc., have been removed. This should be carried out immediately prior to the fitting and securing of panels, covers, etc., as appropriate. The circuits should then be proved, as far as the installation permits, by making ground functioning checks of the services concerned.

A dated record of all relevant figures obtained during the checks should be retained.

Any disconnections or disturbance of circuits associated with flying or engine controls, will require duplicate inspection and functioning tests .

**BONDING**

Bonding is the electrical interconnection of metallic parts of an aircraft normally at earth potential for the safe distribution of electrical charges and currents.

**Function of Bonding.**

Bonding provides a means of protection against charges as a result of the build-up of precipitation static and electrostatic induction as a result of lightning strikes so that the safety of the aircraft or its occupants is not endangered. The means provided are such as to (a) minimise damage to the aircraft structure or components, (b) prevent the passage of such electrical currents as would cause dangerous malfunctioning of the aircraft or its equipment, and (c) prevent the occurrence of high potential differences within the aircraft. Bonding also reduces the possibility of electric shock from the electrical supply system, reduces interference with the functioning of essential services (e.g radio communications and navigational aids) and provides a low resistance electrical return path for electric current in earth-return systems.

**Primary and Secondary Conductors.**

Primary conductors are those required to carry lightning discharge currents, whilst secondary conductors are provided for other forms of bonding. Chapter D4-6 of British Civil Airworthiness Requirements (BCAR) specifies that :-

a. i) The cross-sectional area of primary conductors made form copper must not be less than 0.0045 sq. in (i.e. 1/4 inch by 26 s.w.g.) except that where a single conductor is likely to carry the whole discharge from an isolated section, the cross-sectional area must not be less than 0.009 sq.in. (i.e. 1/2 inch by 26 s.w.g). Aluminium conductors must have a cross-sectional area giving an equivalent surge carrying capacity.

   ii) The cross-sectional area of secondary conductors made from copper must not be less than 0.001 sq. in. which corresponds to 44 strands of 39 s.w.g for braided conductors. Where a single wire is used its size must be not less than 18 s.w.g.

b. BCAR 23 ACB 23.867 and JAR-25 ACJ 25 x 899 (4.2)

   i. Where additional conductors are required to provide or supplement the inherent primary bonding paths provided by the structure or equipment, then the cross sectional area of such primary conductors made from copper should not be less than 3 $mm^2$ except that, where a single conductor is likely to carry the whole discharge from an isolated section, the cross sectional area would be not less than 6 $mm^2$. Aluminium primary conductors should have a cross sectional area giving an equivalent surge carrying capacity.

   ii. Where additional conductors are required to provide or supplement the inherent secondary bonding paths provided by the structure or equipment, should be not less than 1 $mm^2$. Where a single wire is used its size should be not less than 1.2 mm dia.

**Bonding of Aircraft of Metallic and Non-Metallic Construction.**

The skin of an all-metal aircraft is considered adequate to ensure protection against lightning discharge provided that the method of construction is such that it produces satisfactory electrical contact at the joints.

**NOTE :** An electrical contact with a resistance less than 0.05 ohm is considered satisfactory.

With regard to aircraft of nonmetallic or composite construction, a cage, consisting of metallic conductors having a surge carrying capacity at least equal to that required for primary conductors and to which metal parts are bonded, forms part of the configuration of the structure and must conform to the requirements of Chapter D4-6 of BCAR

The earth system which in the case of aircraft of metallic construction is normally the aircraft structure and for aircraft of nonmetallic construction is the complete bonding system, must be automatically connected to ground on landing. This is normally achieved through the nose or tail wheel tyre, which is impregnated with an electrically conducting compound, to provide a low resistance path.

**NOTE :** On some aircraft, a static discharge wick or similar device trailed from a landing gear assembly is used to provide ground contact on landing.

The reduction or removal of electrostatic charges which build up on such surfaces as glass fibre reinforced plastic, can be achieved by the application of a paint, e.g. PR 934, which produces a conductive surface.

**Bonding Connections.**

When a bonding connection is to be made or renewed, it is essential that the conductor has the specified current-carrying capacity, since the bond may have been designed to carry relatively high electrical loads, e.g. under circuit fault conditions.

The manufacturers of solid bonding strip and braided bonding cord usually quote the cross-sectional area on the relevant data sheet. However, in the case of renewal or repair, if the original conductor cannot be matched exactly, a replacement manufactured of the same type of material, but of greater cross-sectional area, should be selected.

Braided copper or aluminium cords fitted at each end with connecting tags or lugs (usually referred to as "bonding jumpers") should be used for bonding connections between moving parts or parts subjected to vibration, and these are suitable both as primary and secondary conductors.

i) The tags or lugs on bonding jumpers are generally fitted by the "crimping method" and only the correct form of crimp and crimping tools should be used for the particular connection. During assembly of the connections to aluminium cords, antioxidant (crimping) compound consisting of 50 per cent by weight of zinc oxide in white petroleum jelly, and complying with DTD 5503, should be applied to the connections.

ii) Where applicable, the soldering of tags or lugs fitted to braided copper cord, using a resin flux. Special care is necessary because overheating and cooling of conductors will cause brittleness, whilst a loss of flexibility up to 1 inch from the lug may occur as a result of the capillary action of the molten solder.

**NOTE:** primary flexible conductors are often made of 600 strands of copper wire, 0.0048 inch in diameter, and formed in a flat braid approximately 0.0625 inch wide.

All bonding connections should be properly locked to prevent intermittent contact which may be caused by vibration.

**NOTE:** Intermittent contact is worse than no contact at all.

Bonding connections should not interfere mechanically or electrically with any associated or adjacent equipment, and bonding jumpers should not be excessively tight or slack.

The run of all primary conductors should be as straight as possible: sharp bends must be avoided.

The number and location of bonding connections to the various components is important and this should be checked and verified by reference to the relevant drawing, e.g. where an engine is not in direct electrical contact with its mounting it should be bonded with at least two primary conductors, one on each side of the engine.

In most instances the following joints are considered self-bonding, provided that all insulting materials (e.g. anodic finish, paint, storage compounds, etc.), are removed from the contact faces before assembly, but if any doubt exists regarding the correctness of the bonds, bonding test should be carried out.

i. Metal-to-metal joints held together by threaded devices, riveted joints, structural wires under appreciable tension and bolted or clamped fittings.
ii. Most cowling fasteners, locking and latching mechanisms.
iii. Metal-to-metal hinges for doors and panels and metal-to-metal bearings (including ball bearings).
    a. In the case of bearings for control surface hinges it should be ascertained which bearings are classified as self bonding, e.g. metal-to-metal, nylon with conducting grease.
    b. Where applicable, bonding jumpers for control surfaces should be as flexible and as short as possible, of as low impedence as is practicable and should not be tinned. The possibility of a jumper jamming the controls must be avoided.

Flexible hose connections used for joining rigid pipes should be bonded by fitting clips around the pipes approximately 1/2 inch away from the hose, and bridging with a corrugated bonding strip or jumper; the practice of tucking the ends of bonding strips between the hose and the pipe is not recommended. To obtain good electrical contact the area under each clip should be cleaned and, after the clip has been fitted, protection should be restored.

i. Not only must the flexible hose connection be bridged, but each pipe run should be bonded to earth at each end, particularly within a radius of 8 feet of any unscreened radio equipment or aerial lead, where earthing bonds should not be more than 5 feet apart, or less distance apart, if called for by the manufacturer.
ii. If bridging strips or bonding cords are fractured a new conductor should be fitted. The soldering of broken ends is prohibited.
iii. High-pressure flexible pipe assemblies are usually self-bonding, but a bonding test should be made between the assembly end-couplings to prove the integrity of the bonding.

**NOTE :** The Provisions of paragraph (i) above also apply to any long electrically-conducting parts (including metallic conduits and metal braiding) when are not insulated from earth.

When any bonding or earth connection is made to the structure or equipment, the specified standard of protection against corrosion should be provided.

i) After a non-conducting protective coating has been removed from the connecting area, the preferred sealing and antioxidants treatment as specified on the relevant drawing and specification should be carried out.

**NOTE :** Non-conducting protective treatments include all generally used priming and finishing paints, varnishes and temporary protective, chromic, anodic and phosphate coatings. Metallic coatings, such as cadmium and tin, are satisfactory conductors, and should not be removed. If a polysulphide compound is used for sealing the earth or bonding point, it must be ensured that the antioxidant to be subsequently applied will not have a detrimental effect on the sealing; e.g. DTD 5503 should not be used.

ii) When the connection has been made any excess compound should be wiped off, using a rag damped in methyl ketone, and the connection and adjacent area re-protected by the specified method, this depending on the materials concerned and the position of the connection.

iii) When a "corrosion washer" forms part of the connecting assembly, it should be correctly fitted and be of the correct material for the type of connection concerned.

**NOTE :** A corrosion washer is plated, or manufactured of material having a potential such that when placed between materials of widely differing potential it reduces the risk of corrosion caused by electrolytic action.

**Earth Terminals.**

When earth-return terminal assemblies are fitted or replaced the correct method of fitting to the structure, the corrosion protection required and the exact location on the structure should be carefully checked. The procedure for fitting and the number of terminations to be attached will vary with the design of the terminal assembly and the type of structure, therefore, reference should be made to the relevant drawings and instructions to ensure both electrical and structural integrity.

i) All earth terminal assemblies should be checked for resistance between the lug attachment point(s) and the surrounding structure and this must not exceed the figure specified for the aircraft concerned (e.g. 0.025 ohm). When earth terminal assemblies are also used to carry electrical supplies, a millivolt drop test, must be carried out

ii) If the resistance in either case is unsatisfactory, the terminal assembly should be removed, the contacting faces cleaned with a fine abrasive (e.g. aluminium wool), and reassembled using, where applicable, new corrosion washers. The connecting area should be sealed and treated with antioxidant compound as specified in the relevant drawing and specification.

**NOTE :** Leads connected to earth terminal assemblies should be of insulated cable with terminal tags fitted by the crimping method. It is important that the cable is of the specified gauge for the service concerned and is kept as short as possible.

**Resistance Values**

Chapter D4-6 of BCAR prescribes the CAA's Requirements with regard to the maximum resistance values for the various conditions of bonding. Those parts of the Requirements which are relevant to the contents of this Leaflet are summarised in Table 12.1.

**TABLE 12.1**

| Bonding Classification | Test Condition | Maximum Resistance |
|---|---|---|
| Primary | Between extremities of the fixed portions of aircraft of non-metallic or composite construction. | Estimated and declared by manufacturer. |
| | Between extremities of the fixed portions of metallic aircraft. | 0.05 ohm |
| | Between bonded components and portions of main earth system to which they are connected. | |
| Secondary | Between metallic parts normally in contact with flammable fluids and main earth system, and also between the parts themselves. | 1 ohm |
| | Between all isolated conducting parts which may be Subjected to appreciable electrostatic charging and the main earth system. | 0.5 megohm or 100,000 ohms per sq. Ft. Of surface area whichever is the less |
| | Between equipment supplied from an unearthed system, of any voltage, and the main earth system. | 1 ohm (See Note 1) |
| | Between equipment containing circuits carrying 50 volts (r.m.s. Or d.c.) Or more, and the main earth system. | |

**NOTES :**

1. The value of 1 ohm is chosen to allow for the inclusion of the resistance of any cable that may be employed for this bonding case, but no one contact resistance should exceed 0.05 ohm.
2. The parts concerned are those situated inside and outside an aircraft and having an area greater than 3 sq. in and a linear dimension greater than 3 inch.

***Bonding Carrying the Main Electrical Supply***

i) The cross-sectional area of the main earth system, or any connection to it, must be such that without overheating or causing excessive voltage drop, it will carry any electrical currents which may pass through it normally or under fault conditions.

ii) If, under fault conditions, it should form part of a short-circuit, not provided against by a protective device, , it should be capable of carrying the full short-circuit current which can pass, without risk of fire or damage to the bonding system.

**NOTE :** For example, paragraph (ii) may apply to bonding which under fault conditions becomes part of a starter or other heavy current circuit. Particular attention should be given to nonmetallic aircraft fitted with a double-pole wiring system to which single-pole equipment has subsequently been added.

**Bond Testing**

Special test equipment, comprising a meter and two cables each of specific length, is required for checking the resistance of bonding. A meter widely used, consists of an ohmmeter operating on the current ratio principle , and a single 1-2 volt nickel-alkaline cell housed in a wooden carrying case. The associated cables are 60 feet and 6 feet in length, and are fitted with a single-spike probe and a double-spike probe respectively. Plug and socket connectors provide for quick-action connection of the cables to the instrument.

Prior to carrying out a bonding test , check should be made on the state of the nickel-alkaline cell of the tester by observing that a full-scale deflection of the meter pointer is obtained when the two spikes of the 6-feet cable probe are shorted. A check should also be made to ensure that when the two spikes are shorted by the single-spike probe of the 60-feet cable, the meter reads zero

The 60-feet lead of the test equipment should be connected to the main earth (also known as the bond datum point) at the terminal points which are usually shown diagrammatically in the relevant aircraft Maintenance Manual. Since the length of a standard bonding tester lead is 60 feet, the measurement, between the extremities of the larger types of aircraft may have to be done by selecting one or more main earth points successively, in which event the resistance value between the main earth points chosen should be checked before proceeding to check the remote point.

**NOTE :** When connecting the 60-feet lead to an earthing point, any protective treatment (e.g. strippable lacquer) should be removed at the point of contact.

The 6-feet test lead should be used to check the resistance between selected points: these are usually specified in the bonding test schedule or the Maintenance Manual for the aircraft concerned. When the two spikes of the test lead probe are brought into contact with the aircraft part, the test-meter will indicate, in ohms, the resistance of the bond.

To ensure good electrical contact at the probe spikes, it may be necessary to penetrate or remove a small area of a non-conducting protective coating. Therefore, after test, any damage to the protective coating must be restored.

If the resistance at a bond connection is excessive, rectification action will depend on the type of connection. The following action should be taken for the more common types of connections.

**NOTE :** Corrosion tends to form at a bonding or earth connection and is often the cause of excessive resistance.

i) In the case of bonding jumpers, the connecting tag or lug should be removed and the contacting faces thoroughly cleaned, using a slight abrasive if necessary. The bare metal thus exposed should be only just large enough to accept the palm of the tag or lug. The connecting area should be sealed and treated with antioxidant as specified in the relevant drawing and specification.

**NOTE :** When an abrasive has been used it is important to ensure that all traces of its are removed.

ii) Where equipment is bonded through a holding bolt, the bolt should be removed and the area under the bolt-head, or nut, thoroughly cleaned and protected as recommended . The correct washer (both with regard to size and material) should be fitted before the bolt is replaced and tightened.

iii) Where the required bond value cannot be obtained at a structural joint the advice of the manufacturer should be sought.

The resistance between the main earth system and a metal plate on which the earthing device (e.g. tyre) is resting should be measured and should not exceed 10 mega ohms when measured with a 250-volt or 500-volt resistance tester, as specified in the test schedule.

**NOTE :-** After carrying out tests, all areas where the protective coating has been removed should be re-protected using the appropriate scheme.

**Bonding Tester Servicing**
A tester requires little in the way of servicing, apart from periodic attention to the alkaline cell, which should be removed at prescribed intervals for routine servicing. When replacing the cell, it is most important that the polarity of connection is correct. The ohmmeter is normally sealed in its case and no attempt should be made to open it: if a fault should develop then the complete instrument should be withdrawn from use and overhauled.

The leads are an integral part of the tester, and being carefully matched to the meter unit must not be modified or altered in any way. All contact surfaces of plug pins and probes must be kept scrupulously clean, and the points of the probe spikes should be reasonably sharp to give effective penetration of protective finishes, etc., on metal surfaces.

The Accuracy of the tester should be checked periodically by using it to measure the resistance of standard test resistors. Normally, three such resistors are supplied for testing purposes and the readings obtained should be within 10% of the standard ohmic values.

**Inspection and testing of Circuits**
***Inspection of Wiring System***
Before carrying out tests, or when inspection is specified in the approved Maintenance schedule, all aircraft circuits, together with plugs, sockets, terminal blocks and equipment terminals, should be examined, as appropriate, for signs of damage, deterioration, chafing, poor workmanship and security of attachments and connections. It is not intended for the purpose of this examination, that electrical apparatus should be removed from its mountings or that cables should be unduly disturbed, but if modifications or repairs, for example, have been carried out in the vicinity, looms should be closely inspected for ingress of metallic swarf between cables. Whenever a structure is opened over wiring which is not normally visible through available inspection panels, circuits so exposed should be thoroughly inspected.

The primary purpose of the inspection is to determine the physical state of the wiring system, especially at bends, points of support, duct entries, etc., or where high temperature or contamination could cause local deterioration. Where cables are grouped together, the state of the outer cables is generally indicative of the condition of the remainder.

Cables completely enclosed in ducts obviously cannot be examined along their length, but should be checked for continuity and insulation, especially if oil or water ingress is suspected. Where there is evidence of damage to the ducts, the cables should be exposed to ascertain their condition.

Terminations must be secure and good electrical contact obtained without strain on the threads of terminal pillars or studs. Torque loadings, where appropriate, should be within the limits specified.

**CONTINUITY TESTING**
A concealed break in a cable core or at a connection may be found by using a continuity tester which normally consists of a low voltage battery (2.5 volts is satisfactory) and a test lamp or low reading voltmeter.

**NOTE :-** In some testers incorporating a test lamp, semiconductors are included in the test lamp circuit and, to prevent damage, the currents should be limited to 120 milli amps.

Before testing, the main electrical supply should be switched off or disconnected. A check should be made that all fuses are intact and that the circuit to be tested is not disconnected at any intermediate point. All switches and circuit breakers, as appropriate, should be closed to complete the circuit.

When carrying out a low voltage continuity check, it is essential to work progressively through the circuit, commencing from the relevant fuse or circuit breaker and terminating at the equipment. Large circuits will probably have several parallel paths and these should be progressed systematically, breaking down as little as possible at plug and socket or terminal block connection. In testing of this nature, it is valueless to check several low resistance paths in parallel.

**MILLIVOLT DROP TEST**
Excessive resistance in high-current carrying circuits can be caused by loose terminal connections, poorly swaged lead ends, etc. Faults of this kind are indicated by low terminal voltage at the connections to the service load and by heating at a conductor joint. If such faults are suspected, a millivolt drop test as described below is recommended, but it is also acceptable to check along progressive sections of the system with an accurately calibrated voltmeter.

i) For continuously-rated circuits, the test should, whenever possible, be made with the normal operating current flowing, the power being derived from an external source. For short-rated circuits, a suitable resistance or other dummy load should be used in lieu of the normal load and the current should be scaled down to avoid overheating.

**NOTE :-** The test voltage may be reduced for safety reasons.

ii) The millivolt-meter should be connected to each side of the suspected joint and a note made of the volt drop indicated. The indicated reading should be compared with the figures quoted in the relevant publication (an approximate guide is 5 mV/10 amps. flowing).

**INSULATION RESISTANCE TESTING**

In the following paragraphs general test procedures are outlined; however, as a result of the wide variation in electrical installation and equipment which exists with different aircraft, the routing charts and approved test schedule for the aircraft concerned must be consulted. All ancillary equipment should be tested separately in accordance with the appropriate manufacturers' publications.

After installation and where specified in the approved Maintenance Schedule or test schedule, aircraft circuits should be tested by means of a 250-volt insulation tester which should have its output controlled so that the testing voltage cannot exceed 300 volts. In all systems having nominal voltages over 30 volts, cables forming circuits essential to the safety of the aircraft should be tested individually. Other circuits may be connected in groups for test. However, the numbers of circuits which may be grouped for test is governed by the test results; where the insulation resistance so measured is found to be less than the appropriate minimum value stated , the number of circuits grouped together should be reduced.

Immediately after an insulation test, functioning checks should be made on all the services subjected to the test. If the insulation test or subsequent functioning tests should reveal a fault, the fault should be rectified and the insulation and functioning tests should be repeated in that sequence on the affected circuits.

**Preparations Prior to Test**

Before beginning an insulation test on a system, the following preparations should be made, details of which will depend on the installation concerned.

i) The aircraft battery and any external supply should be disconnected.
ii) Where applicable, circuit breakers should be closed.
iii) The power selector switch should be switched to the position appropriate to that required for normal in-flight operation.
iv) All switches in the circuit concerned should be 'ON', dimmer-switches should be set at the minimum resistance position and micro-switches operated to the 'ON' position.
v) All items of ancillary equipment which are supplied by the system concerned should be disconnected. This includes all rotary equipment (e.g. generators, motors, actuator units, etc.), radio equipment, capacitors, semiconductors, voltage regulator coils, electrical instruments, fire extinguishers, etc.
vi) In cases where the insulation resistance with the items connected is not less than 2 mega ohms, the disconnection may be made by the earth lead, leaving the item connected to the circuit.

**NOTE :-** Bonded earth connections to the airframe structure should, if possible, remain undisturbed for the purpose of these tests.

vii) Components such as cut-outs and relays which are normally open should have their terminals bridged to ensure continuity of the circuit, and disconnected leads from suppressors should also be bridged for similar reasons. Where a suppressor cannot be bridged and plug and socket connections are used, the capacitors should be discharged before the circuit is reconnected, otherwise arcing and burning of the pins may occur. Items in series which are disconnected should also be bridged so that part of the circuit is not omitted.

**Testing the System**

i) Double-pole systems on some older types of aircraft can be tested by connecting the leads of the insulation tester to each of the battery leads and measuring the resistance between them and, afterwards, checking the resistance between each battery lead and earth; fuses should be left in position for this test. On some large aircraft with double-pole systems, cables may be grouped as for single-pole systems, the earthing checks being made between punched positive and earth and bunched negative and earth.
ii) To test single-pole systems, on lead of the tester should be connected to earth and the other to the cable or bunch of cables to be tested. When cables are bunched together, it is advisable to limit the number to the smallest convenient figure. If the insulation resistance is less than the appropriate value, the number of circuits should be reduced. Testing should continue until, by process of elimination, any defective cables have been identified

**Test Result.**

The results of insulation tests are of little significance unless they are related to test results obtained on other occasions. The insulation resistance values are likely to vary with changes in the temperature and humidity of the local atmosphere, e.g. if the aircraft has been in damp conditions for some time before the test, low teadings can be expected. Results of tests and the temperature and humidity conditions at the time of the test should be recorded, so that any pronounced drop in resistance found on subsequent tests can be checked and rectified as necessary.

Section J of British Civil Airworthiness Requirements does not specify minimum values of insulation resistance, but gives guidance on values that may be expected during maintenance testing. These values can be, and frequently are, exceeded considerably on new installations. The values given are as follows:-

i) Wiring (including accessories for jointing and terminating ):-
In engine nacelles, undercarriage wheel wells and other situations exposed to weather or extremes of temperature .... 2 mega ohms

Galley and other nonessential services, lighting, signalling and indication services .. .... 5 mega ohms

Other services .. .. .. .... 10 mega ohms

ii) Wiring accessories alone (e.g. terminal blocks, connectors, plugs and sockets, etc.):-
Between terminals .. .. .... 100 mega ohms
Between terminals bunched together and earth 200/number of terminals mega ohms

iii) Rotating machinery whichever is the greater of rated voltage/150 or 0.5 mega ohms

iv) All other equipment (including indicating instruments) .. ... 5 mega ohms

**FUNCTIONING TESTS**

Before conducting any tests, all precautions for aircraft and personnel safety should be taken. Whenever possible, functioning tests should be carried out using an external supply coupled to the ground supply connector. Tests must ensure proper functioning of individual and integrated sections of circuits, and should be in accordance with schedules established by reference to details in the relevant Maintenance Manual, Wiring Diagram Manual or, where appropriate, instructions relating to the incorporation of a modification or any substantial rewiring .

**NOTE :-** Where applicable, when one or more engines are running, the power supply can be obtained from the associated generators, due reference being made to the functioning of any isolating relays.

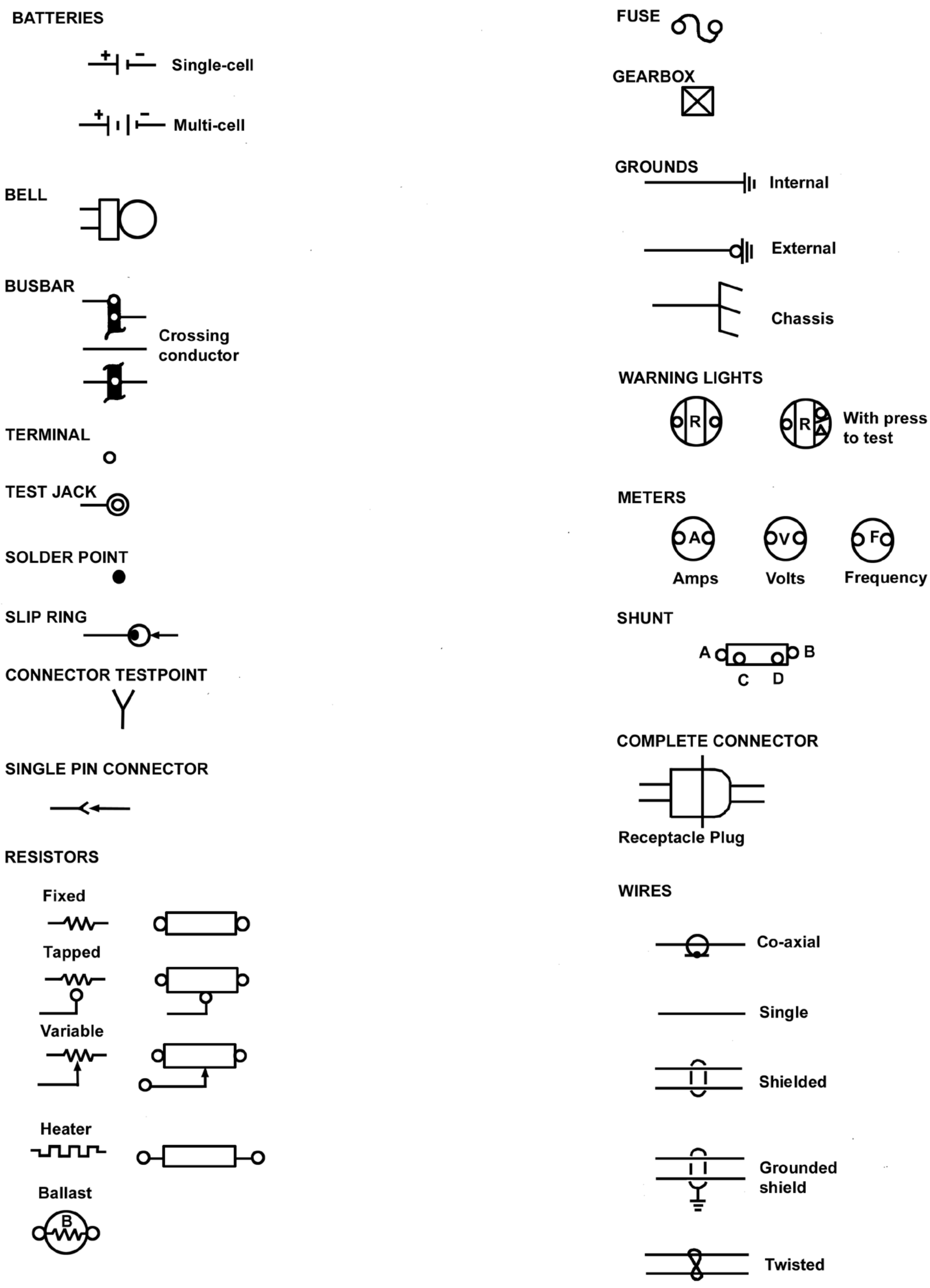
For certain circuits (e.g. stand-by lighting), functioning tests can only be carried out using the aircraft battery system, but this battery should be used as little as possible.

After the normal functioning test of an individual circuit has been completed and the circuit switched off, the fuse should be removed or the circuit breaker tripped and the circuit again switched on to check the isolation of the circuit concerned.

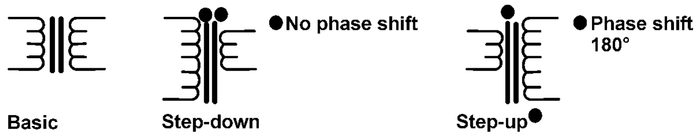
When the operation of a circuit (e.g. generator equaliser circuit) depends on the inherent resistance value of the circuit, the resistance should be measured with a low reading ohmmeter (such as that used in a bonding tester) to determine that the resistance is within the specified limits.

■■■

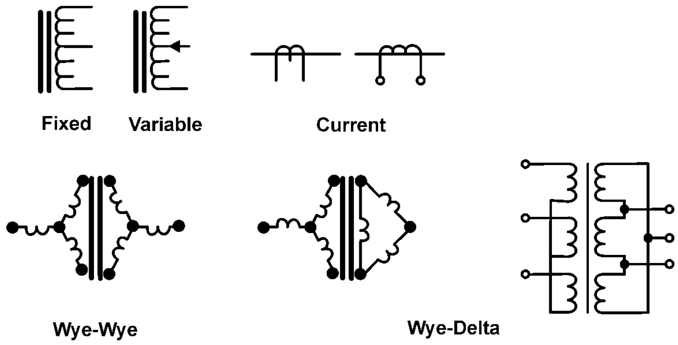
# CHAPTER : 13
# ELECTRICAL DIAGRAM SYMBOLS



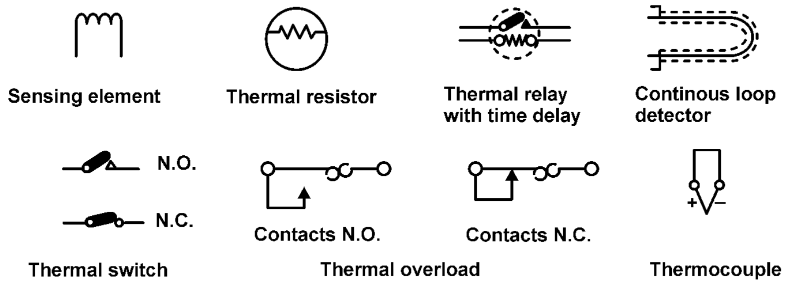
TRANSFORMERS



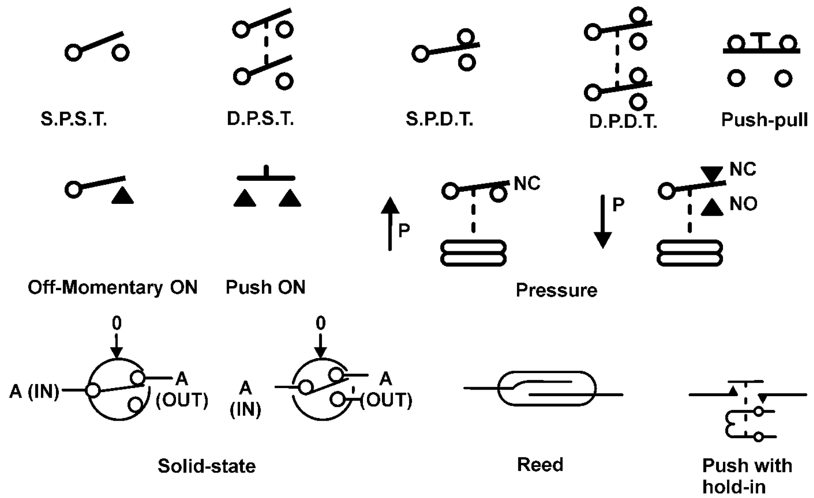

AUTO



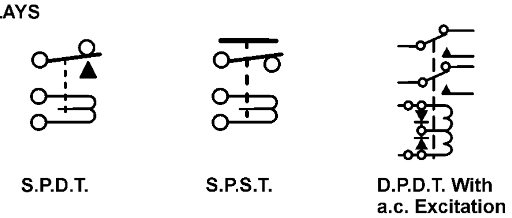

THERMAL DEVICES




SWITCHES




RELAYS



# INSTRUMENT SYSTEM

# CHAPTER : 14
# AIRCRAFT INSTRUMENT PANELS AND RANGE MARKING

## GENERAL

Safe, economical, and reliable operation of modern aircraft is dependent upon the use of instruments. The first aircraft instruments were fuel and oil pressure instruments to warn of engine trouble so that the aircraft could be landed before the engine failed. As aircraft that could fly over considerable distances were developed, weather became a problem. Instruments were developed that helped to fly through bad weather conditions.

Instrumentation is basically the science of measurement. Speed, distance, altitude, attitude, direction, temperature, pressure, and r.p.m. are measured and these measurements are displayed on dials in the cockpit.

There are two ways of grouping aircraft instruments. One is according to the job they perform. Within this grouping they can be classed as flight instruments, engine instruments, and navigation instruments. The other method of grouping aircraft instruments is according to the principle on which they work. Some operate in relation to changes in temperature or air pressure and some by fluid pressure. Others are activated by magnetism and electricity, and others depend on gyroscopic action.

The instruments that aid in controlling the in-flight attitude of the aircraft are known as flight instruments. Since these instruments must provide information instantaneously, they are located on the main instrument panel within ready visual reference of the pilot. Basic flight instruments in an aircraft are the airspeed indicator, altimeter and the magnetic direction indicator. In addition, some aircraft may have a rate-of-turn indicator, a bank indicator, and an artificial horizon indicator. Flight instruments are operated by atmospheric, impact, differential, or static pressure or by a gyroscope.

Engine instruments are designed to measure the quantity and pressure of liquids (fuel and oil) and gases (manifold pressure), r.p.m., and temperature. The engine instruments usually include a tachometer, fuel and oil pressure gauges, oil temperature gauge, and a fuel quantity gauge. In addition some aircraft that are powered by reciprocating engines are equipped with manifold pressure gauge (s), cylinder head temperature gauge (s), and carburettor air temperature gauge (s). Gas turbine powered aircraft will have a turbine or tailpipe temperature gauge (s), and may have an exhaust pressure ratio indicator (s).

Navigational instruments provide information that enables the pilot to guide the aircraft accurately along definite courses. This group of instruments includes a clock, compasses (magnetic compass and gyroscopic directional indicator), radios, and other instruments for presenting navigational information to the pilot.

## INSTRUMENT CASES

A typical instrument can be compared to a clock, in that the instrument has a mechanism, or works; a dial, or face; pointers, or hands; and a cover glass. The instrument mechanism is protected by a one; or two-piece case. Various materials, such as aluminum alloy, magnesium alloy, iron, steel, or plastic are used in the manufacture of instrument cases. Bakelite is the most commonly used plastic. Cases for electrically operated instruments are made of iron or steel ; these materials provide a path for stray magnetic force fields that would otherwise interfere with radio and electronic devices.

Some instrument mechanisms are housed in airtight cases, while other cases have a vent hole. The vent allows air pressure inside the instrument case to vary with the aircraft's change in altitude.

## DIALS

Numerals, dial markings, and pointers of instruments are frequently coated with luminous paint. Some instruments are coated with luminous calcium sulphide, a substance that glows for several hours after exposure to light. Other instruments have a phosphor coating that glows only when excited by a small ultraviolet lamp in the cockpit. Some instruments are marked with a combination of radioactive salts, zinc oxide, and shellac. In handling these instruments, care should be taken against radium poisoning. The effects of radium are cumulative and can appear after a long period of continued exposure to small amounts of radiation. Poisoning usually results from touching the mouth or nose after handling instrument dials or radioactive paint. After handling either, the hands should be kept away from the mouth and nose, and washed thoroughly with hot water and soap as soon as possible.

## RANGE MARKINGS

Instrument range markings indicate, at a glance whether a particular system or component is operating in a safe and desirable range of operation or in an unsafe range.

Instruments should be marked and graduated in accordance with the Aircraft Specifications or Type Certificate Data Sheets and the specific aircraft maintenance or flight manual. Instrument markings usually consist of colored decalcomanias or paint applied to the outer edges of the cover glass or over the calibrations on the dial face. The colors generally used as range markings are red, yellow, green, blue, or white. The markings are usually in the form of an arc or a radial line.

A red radial line may be used to indicate maximum and minimum ranges : operations beyond these markings are dangerous and should be avoided. A blue arc marking indicates that operation is permitted under certain conditions. A green arc indicates the normal operating range during continuous operation. Yellow is used to indicate caution.

A white index marker is placed near the bottom of all instruments that have range markings on the cover glass. The index marker is a line extending from the cover glass onto the instrument case. The marker shows if glass slippage has occurred. Glass slippage would cause the rage markings to be in error.

**INSTRUMENT PANELS**

With a few exceptions, instruments are mounted on a panel in the cockpit so that the dials are plainly visible to the pilot or copilot. Instrument panels are usually made of sheet aluminum alloy strong enough to resist flexing. The panels are nonmagnetic and are painted with a nonglare paint to eliminate glare or reflection.

In aircraft equipped with only a few instruments, only one panel is necessary; in some aircraft, additional panels are required. In such cases the forward instrument panel is usually referred to as the "main" instrument panel to distinguish it from additional panels on the cockpit overhead or along the side of the flight compartment. On some aircraft the main instrument panel is also referred to as the pilot's or copilot's panel, since many of the pilot's instruments on the left side of the panel are duplicated on the right side.

The method of mounting instruments on their respective panels depends on design of the instrument case. In one design, the bezel is flanged in such a manner that the instrument can be flush-mounted in its cutout from the rear of the panel. Integral self-locking nuts are provided at the rear faces of the flange corners to receive mounting screws from the front of the panel. The flanged type case can also be mounted from the front of the panel.

The mounting of instruments that have flangeless cases is a simpler process. The flangeless case is mounted from the front of the panel. A special expanding type of clamp, shaped and dimensioned to fit the instrument case, is secured to the rear face of the panel. A actuating screw is connected to the clamp and is accessible from the front of the panel. The screw can be rotated to loosen the clamp, permitting the instrument to slide freely into the clamp. After the instrument is positioned, the screw is rotated to tighten the clamp around the instrument case.

Instrument panels are usually shock-mounted to absorb low-frequency, high-amplitude shocks. Shock mounts are used in sets of two, each secured to separate brackets. The two mounts absorb most of the vertical and horizontal vibration, but permit the instruments to operate under conditions of minor vibration. A cross sectional view of a typical shock mount is shown in figure 14.1.

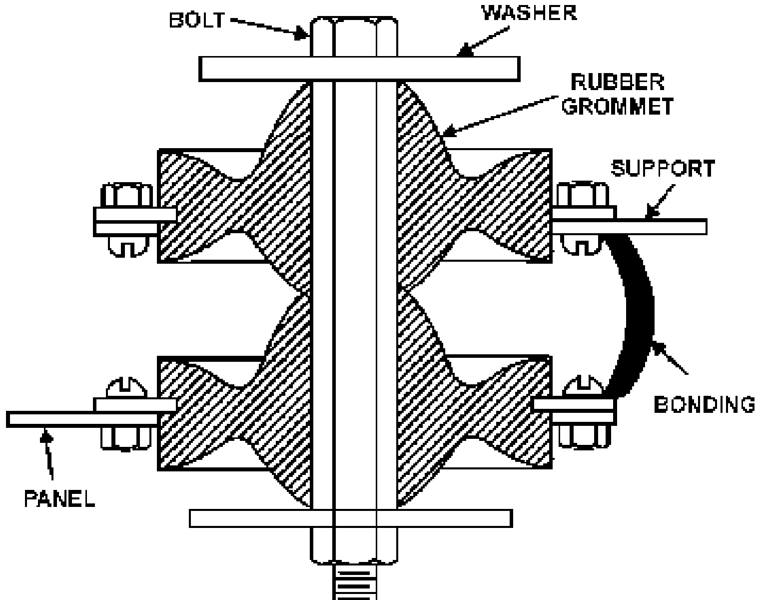



*Fig. 14.1, Section through instrument panel shock.*

The type and number of shock mounts to be used for instrument panels are determined by the weight of the unit. The weight of the complete unit is divided by the number of suspension points. For example, an instrument panel weighing 16 lbs. which is supported at four points would require eight shock absorbers, each capable of supporting 4 lbs. When the panel is mounted, the weight should deflect the shock absorbers approximately 1/8 in.

Shock-mounted instrument panels should be free to move in all directions and have sufficient clearance to avoid striking the supporting structure. When a panel does not have adequate clearance, inspect the shock mounts for looseness, cracks, or deterioration.

■■■

# CHAPTER : 15
# FLIGHT INSTRUMENTS : PITOT-STATIC SYSTEMS

**SOURCES OF PITOT AND STATIC PRESSURES**

Pitot pressure, which varies with airspeed and air density, is the ram air pressure built up by the movement of an aircraft through the air and is sensed by a pressure head externally located at some accurately selected position. Static pressure is the ambient pressure prevailing at the altitude at which the aircraft is flying and may also be sensed by pitot-static head or, as in some aircraft, by a separate static vent system. Both pressures are communicated to the flight instruments through pipelines; the airspeed indicator and other airspeed measuring instruments utilise pitot pressure and static pressure, while instruments such as the altimeter and the vertical speed (rate-of-climb) indicator utilise static pressure only.

**Pressure Heads**

Pressure heads are of two main types, i.e., the pitot-static type which senses and transmits both pitot and static pressures and the pitot-only type which is used in conjunction with a separate static vent system.

Pressure head pipe connections are arranged to suit the method of installation, i.e. frontal installation, fuselage side, or under-wing installation. In the frontal installation the connections emerge in line with the head while in the other two they emerge at 90 degree to the axis of the head. The connections may be of the low pressure type for use with rubber grommets, or of the high pressure type for use with flared pipes and collets.

As prevention against ice formation, pressure heads are equipped with an internal system of appropriately positioned heating elements. Pressure heads designed for fuselage side or under wing mounting have additional elements located inside the support mast. In many cases separate static vents also incorporate heating elements. The heaters are supplied with power from the aircraft electrical system and controlled from the cockpit. Display panels for open circuit detection of elements may also be fitted.

**Installation**

Pressure heads should be examined for physical damage and freedom from obstruction, including drain holes, before installation, and it should be confirmed that the type of head is correct for the particular aircraft. Before connecting heater element cables it must be ensured that the pitot heating circuit of the aircraft is isolated from the electrical power sources.

Drain holes of a calibrated size are provided in the bodies of some types of electrically-heated pressure heads and these must be at the bottom when such heads are installed. Certain types of pressure head designed for mounting on the sides of a fuselage are "handed", and the drain holes and support mast drain screw are located in such a way that they are at the lowest point compatible with the angular position of the head when installed. It is therefore important to ensure that the correctly handed pressure head is fitted.

In order to prevent distortion of mounting flanges, fixing screws or bolts, pressure heads must be adequately supported during installation and not allowed to hang on their fixing screws when these are being tightened.

Pressure heads should not be painted as this may impair their thermal efficiency. Furthermore, paint may cause inadvertent obstruction of the necessary orifices and result in inaccurate sensing of pitot and static pressures.

After installation, the pressure connection should be checked for security and locking, and a leak test of the complete pilot-static system carried out in accordance with the requirements of the aircraft Maintenance Manual. The heating elements of electrically heated pressure heads should be checked for functioning by connecting an ammeter in the circuit and noting that the current consumed is correct for the voltage and power ratings of the pressure head installed. As pitot heads are in positions which make them vulnerable to lightning strikes, bounding should be checked for effectiveness.

**Static Vents**

In order to minimise the effects of position error and to provided greater freedom from ice formation, the sensing of static pressure is by means of separate vents in the forms of flat metal plates secured to the fuselage skin at predetermined positions. There are two principal types of static vent in use, their application being governed by the size and the number of pitot-static system required for a specific aircraft type.

In the basic system, the static vent consists of a flat brass plate, rounded at the ends and having through its centre a 6 mm (0.25 in) diameter hole communicating with a short section of plain pipe which provides for the connection of the vent with the pipeline system. The pipe section may, in some versions, be positioned at 90 degree to the plate or directed upward at 30 degree to provide drainage for moisture. A drain tap is usually fitted in the section of pipeline immediately adjacent to vents of the former type.

In aircraft employing several independent pitot-static systems, the static vent consists of a flat stainless steel plate through which are drilled a number of 5 mm (0.188 in) diameter holes. Each hole is connected to the pipeline of a specific system or component requiring static pressure, by means of a threaded coupling adapter welded concentric with the hole. Holes in the fuselage skin accommodate the coupling adapters of the plate which is bolted to the skin.

**Installation**

The vents are normally fitted in a position free of turbulence from aerials or other external fitments and located where the skin in the area of the vent is flush riveted and free of butt straps, etc., since such features would cause a varying behaviour of the boundary layer.

In order to reduce errors due to pressure imbalance at the vents whenever yawing of the aircraft takes place, static vents are fitted on each side of the fuselage and are interconnected into the same static pressure line.

Brass static vent plates should be provided with a stiffener plate on the inside of the fuselage skin to prevent distortion of the vent plate. On metal aircraft, a metal stiffener, similar in shape to the vent plate, should be riveted to the skin, while with aircraft having a plywood skin, a plywood stiffener should be glued to the skin.The vent plates should make complete contact at their outer edges and be secured to the fuselage skin by means of the spigots provided on the rear face, and by knurled nuts which should be tightened by hand and wirelocked; overtightening may cause the vent plate to distort. A suitable sealing compound should be used between the contracting surfaces of the vent and skin. Exuded sealant should be trimmed off by using a plastic scraper.

Stainless steel vent plates should be bolted directly to the fuselage skin and, depending on the modification state of the aircraft system, are sealed either by sealing rings around the flanges of the coupling adapters, or by filling the inner surface recess with sealing compound.

Smoothness of the outer surface of all static vent plates is vital to the accurate sensing of static pressure. They must therefore be kept free of scratches and other indentations, and must never be painted.

**Pressure Error**

Pressure error may be defined as that part of the difference between the calibrated air speed and the indicated air speed due to the recorded static pressure not being equal to the ambient pressure. The error, which is strongly influenced by the position of pressure heads and static vents, is determined for each type of aircraft by conducting a series of prototype test flights over various ranges of speed, altitude, configuration, weight etc. Details of the error so determined, and the corrections to be applied to the readings of airspeed indicators and altimeters, are presented in either tabular or graphical form and contained in an appropriate section of the aircraft's Flight Manual.

Any subsequent alteration in the position of either pressure heads or static vents, the provisioning of additional systems, or alterations occasioned by modifications or repair to the aircraft in the vicinity of pressure heads or static vents may affect the pressure error, necessitating further test flights and alterations to the Flight Manual.

**Pipelines**

Pitot and static pressures are transmitted throughout systems by means of light alloy pipes (tungum pipes may also be used in some aircraft) and flexible hoses such as nylon 11, nylon 66, or rilsan, the latter being used for the connection of resilient mounted instruments and components. In order to prevent moisture blockage and to minimise pressure lag, the inside diameter of pipelines must not less than 6 mm (0.25 in).

The procedure for the installation and removal of pipelines depends upon the size and complexity of individual systems; reference should always be made to the aircraft Maintenance Manual. The following points, common to applied practices, are given as a general guide:

a) Before installation, pipelines should be blown out with a clean, dry low-pressure air supply to ensure cleanliness and freedom from obstruction.
b) When tightening the end connections of flexible hoses it must be ensured that the hoses do not become twisted; a yellow or white tell tale line running along the length of the hose will indicate any twisting.
c) Bending of pipes through too small a radius, and kinking, must be avoided since the resulting depressions and reduction in bore diameter will create unwanted moisture traps and erratic transmission of pressure.
d) Metal pipes must be securely attached to the airframe structure at regular intervals throughout their run and should slope towards point at which drain traps or drain valves are located.
e) Pipelines leading from static vents should be installed so that they rise continuously towards the instruments but, if this cannot be achieved, they should rise for the first 150 mm (6 in) at least. Where two static vents are interconnected, the pipelines from each should be symmetrically disposed.
f) Clearance must exist between a newly installed pipe and other pipes or structural parts to avoid chaffing during flight.
g) When pipelines are removed from an aircraft, blanks should be fitted to their end connections and all other connections in the system which become exposed by the removal. Support clamps should be returned to the their original positions as soon as a pipeline has been removed to ensure their correct location.
h) The mating surfaces of pipe ends and connections must be clean.

j) When connecting pipelines employing low-pressure unions and rubber grommets, the union nuts should be tightened by hand and should then be secured by a half-turn with a spanner, since overtightening may damage the grommet and result in a leaking joint.

k) After installation of a pipeline, the system with which it is associated must be checked for leaks in the manner prescribed in the aircraft Maintenance Manual. On satisfactory completion of such individual checks, a leak test of the complete pitot static system must be carried out.

**Drains**

In addition to the moisture drainage of pressure heads, facilities must also be provided for the removal of moisture which might accumulate in the pipelines connecting the sources of pressure to instrument and associated equipment. Such draining facilities may take the form of either drain traps or drain valves (in some aircraft they are used in combination) located at the lowest points in pitot and static pipe runs. The design, construction and application of drains varies between manufacturers and type of pitot-static system. Reference should therefore be made to the Manuals concerning the component and aircraft type.

**PITOT-STATIC SYSTEM**

Three of the most important flight instruments are connected into a Pitot-static system. These instruments are the airspeed indicator, the altimeter, and the rate-of-climb indicator. Figure 15-1 shows these three instruments connected to a Pitot-static tube head.

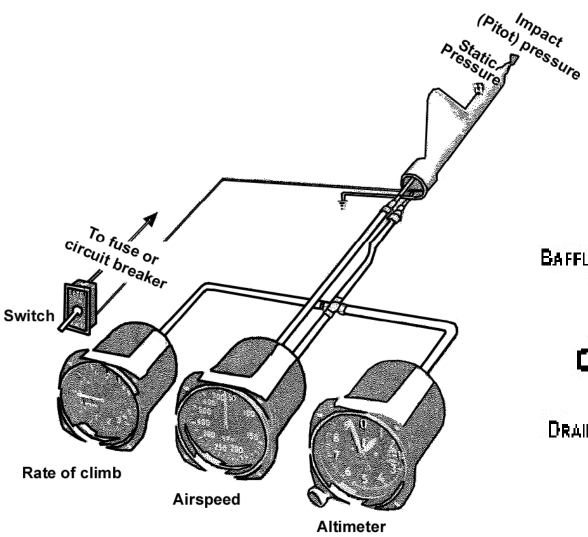



*Fig. 15.1, Pitot-static system.*

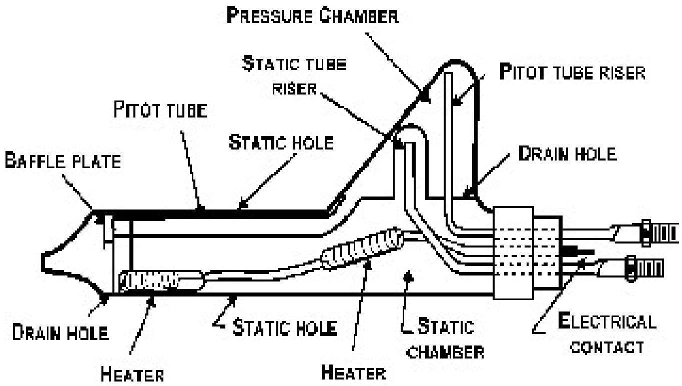



*Fig. 15.2, Pitot-static system head.*

The Pitot-static system head, or Pitot-static tube as it is sometimes called, consists of two sections. As shown in figure 15.2, the forward section is open at the front end to receive the full force of the impact air pressure. At the back of this section is a baffle plate to protect the Pitot tube from moisture and dirt that might otherwise be blown into it. Moisture can escape through a small drain hole at the bottom of the forward section.

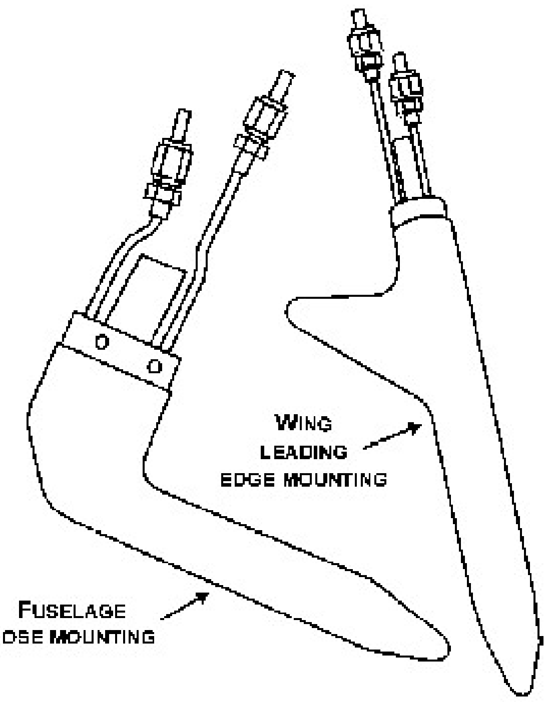
The Pitot, or pressure, tube leads back to a chamber in the "shark-fin" projection near the rear of the assembly A riser, or upright tube, leads the air from this chamber through tubing to the airspeed indicator.

The rear, or static, section of the Pitot-static tube head is pierced by small openings on the top and bottom surfaces. These openings are designed and located so that this part of the system will provide accurate measurements of atmospheric pressure in a static, or still, condition. The static section contains a riser tube which is connected to the airspeed indicator, the altimeter, and the rate-of-climb indicator.

Many Pitot-static tubes are provided with heating elements to prevent icing during flight (figure 15.2). During ice-forming conditions, the electrical heating elements can be turned on by means of a switch in the cockpit. The electrical circuit for the heater element may be connected through the ignition switch. Thus, in case the heater switch is inadvertently left in the "on" position, there will be no drain on the battery when the engine is not operating.

The Pitot-static tube head is mounted on the outside of the aircraft at a point where the air is least likely to be turbulent. It is pointed in a forward direction parallel to the aircraft's line of flight. One general type of tube head is designed for mounting on a streamlined mast extending below the nose of the aircraft fuselage. Another type is designed for installation on a boom extending forward of the leading edge of the wing. Both types are shown in figure 15.3. Although there is a slight difference in their construction, they operate identically.

Most Pitot-static tubes are manufactured with a union connection in both lines from the head, near the point at which the tube head is attached to the mounting boom or mast (figure 15.3). These connections simplify removal and replacement, and are usually reached through an inspection door in the wing or fuselage. When a Pitot-static tube head is to be removed, these connections should be disconnected before any mounting screws and lock-washers are removed.



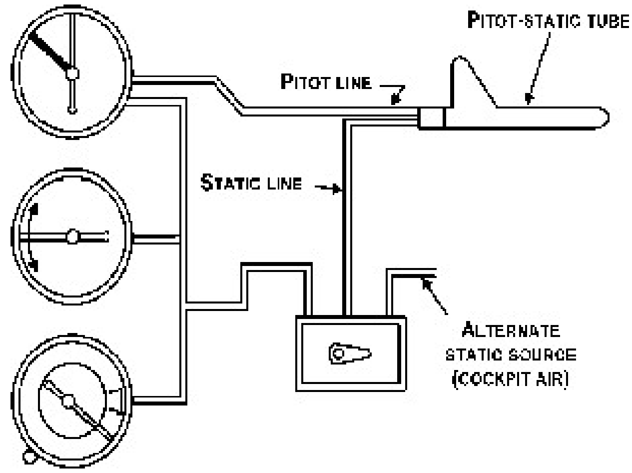




*Fig. 15.3, Pitot-static tube heads.* *Fig. 15.4, Pitot-static system with alternate source of static pressure.*

In many aircraft equipped with a Pitot-static tube, an alternate source of static pressure is provided for emergency use. A schematic diagram of a typical system is shown in figure 15.4. As shown in the diagram, the alternate source of static pressure may be vented to the interior of the aircraft.

Another type of Pitot-static system provides for the location of the Pitot and static sources at separate positions on the aircraft. This type of system is illustrated in figure 15.5.

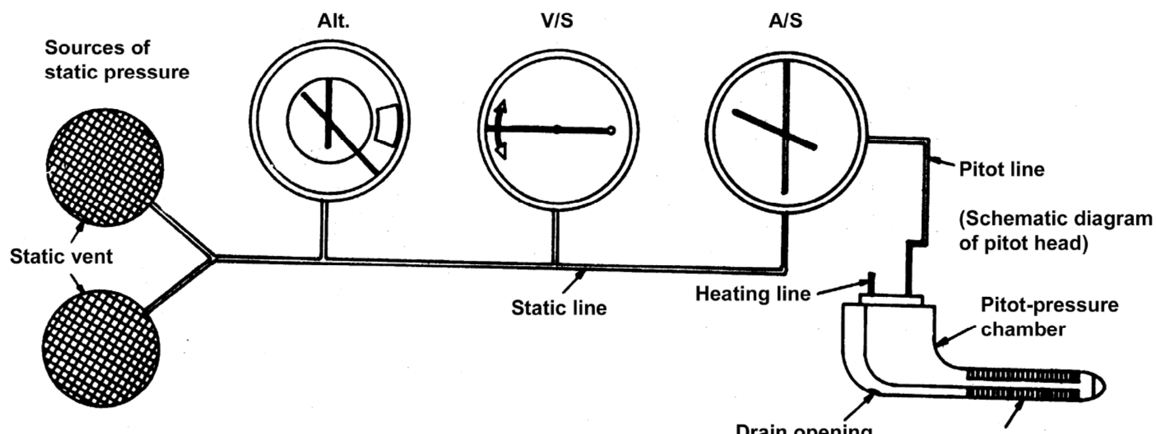



*Fig. 15.5, Pitot-static system with separate sources of pressure.*

Impact pressure is taken form the Pitot-head (figure 15.5) which is mounted parallel to the longitudinal axis of the aircraft and generally in line with the relative wind. The leading edge of the wing, nose section, or vertical stabilizer are the usual mounting positions, since at those points there is usually a minimum disturbance of air due to motion of the aircraft.

Static pressure in this type of Pitot-static system is taken from the static line attached to a vent or vents mounted flush with the fuselage or nose section. On aircraft using a flush-mounted static source, there may be two vents, one on each side of the aircraft. This compensates for any possible variation in static pressure on the vents due to erratic changes in aircraft attitude. The two vents are usually connected by a Y-type fitting. In this type of system, clogging of the Pitot opening by ice or dirt (or failure to remove the Pitot cover ) affects the airspeed indicator only.

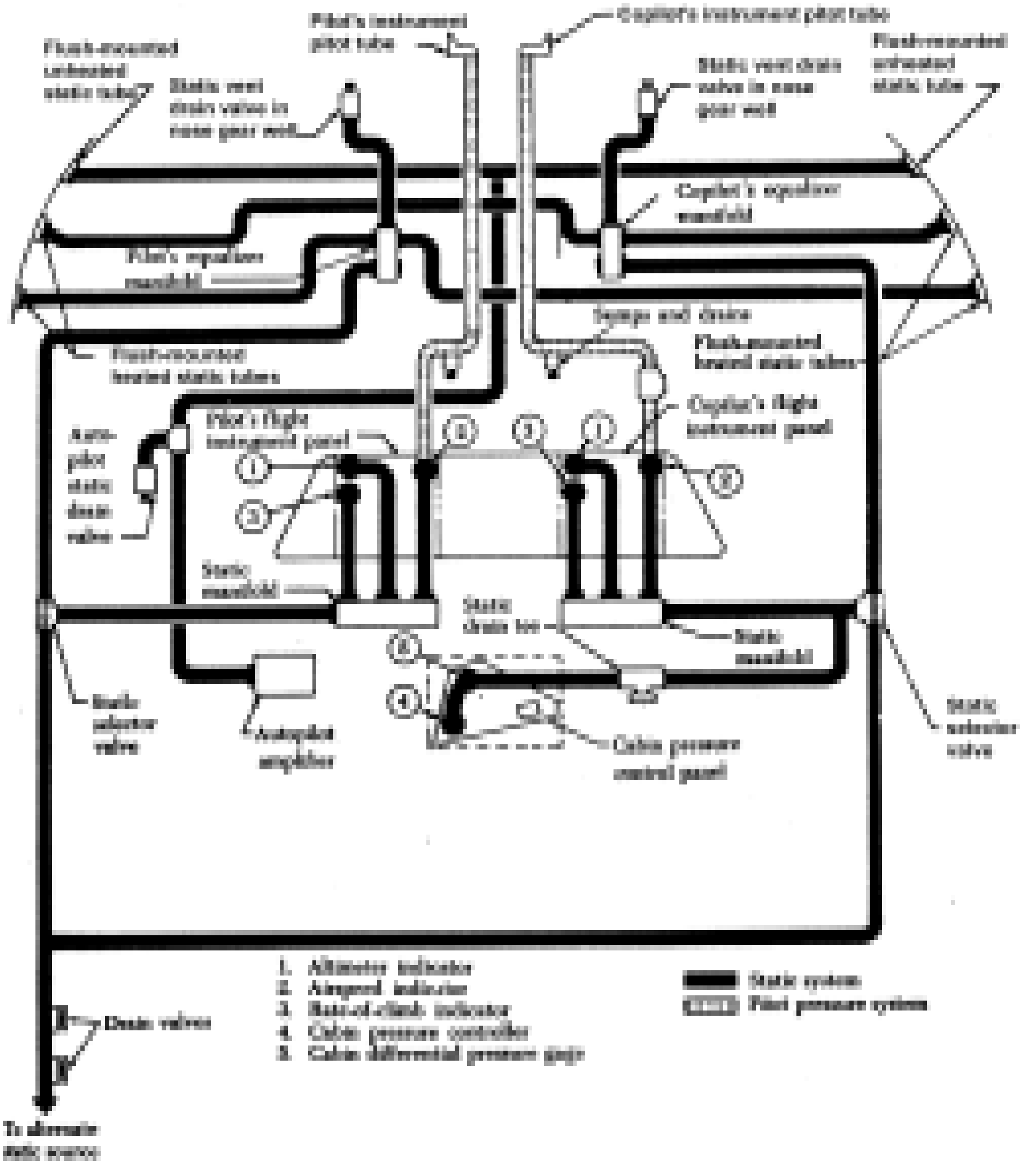


*Fig. 15.6, Schematic of typical Pitot-static system on pressurized multi-engine aircraft.*

A Pitot-static system used on a pressurized, multi-engine aircraft is shown in figure 15.6. Three additional units, the cabin pressure controller, the cabin differential pressure gauge, and the autopilot system are integrated into the static system. Both heated and unheated flushmounted static ports are used.

## MAINTENANCE OF PITOT-STATIC SYSTEMS

The specific maintenance instructions for any Pitot-static system are usually detailed in the applicable aircraft manufacturer's maintenance manual. However, there are certain inspections, procedures, and precautions to be observed that apply to all systems.

Pitot tubes and their supporting masts should be inspected for security of mounting and evidence of damage. Checks should also be made to ensure that electrical connections are secure. The Pitot pressure entry hole, drain holes, and static holes or ports should be inspected to ensure that they are unobstructed. The size of the drain holes and static holes is aerodynamically critical. They must never be cleared of obstruction with tools likely to cause enlargement or burring.

Heating elements should be checked for functioning by ensuring that the Pitot tube begins to warm up when the heater is switched "on". If an ammeter or loadmeter is installed in the circuit, a current reading should be taken.

The inspections to be carried out on the individual instruments are primarily concerned with security, visual defects, and proper functioning. The zero setting of pointers must also be checked. At the time of inspecting the altimeter, the barometric pressure scale should be set to read field barometric pressure. With this pressure set, the instrument should read zero within the tolerances specified for the type installed. No adjustment of any kind can be made, if the reading is not within limits, the instrument must be replaced.

**Leak Testing Pitot-Static Systems**

Aircraft Pitot-static systems must be tested for leaks after the installation of any component parts, when system malfunction is suspected, and at the periods specified in the Regulations.

The method of leak testing and the type of equipment to use depends on the type of aircraft and its Pitot-static system. In all cases, pressure and suction must be applied and released slowly to avoid damage to the instruments. The method of testing consists basically of applying pressure and suction to pressure heads and static vents respectively, using a leak tester and coupling adapters. The rate of leakage should be within the permissible tolerances prescribed for the system. Leak tests also provide a means of checking that the instruments connected to a system are functioning properly. However, a leak test does not serve as a calibration test.

Upon completion of the leak test, be sure that the system is returned to the normal flight configuration. If it was necessary to blank off various portions of a system, check to be sure that all blanking plugs, adapters, or pieces of adhesive tape have been removed.

**ALTIMETERS**

As the name implies, altimeters measure the altitude of the aircraft either above sea-level or above a point of known altitude such as an aerodrome. They are basically sensitive pressure gauges which operate on the aneroid barometer principle and indicate the changes in atmospheric pressure which occur with changes in altitude. The calibration of altimeters is in accordance with the ICAO law for the standard atmosphere, which is based on certain assumed values of pressure and temperature to provide a conventional relationship between pressure and altitude.

An altimeter mechanism consists of a stack of two or in some versions, three aneroid capsules which are exhausted of air and sealed, and connected to a pointer mechanism via a lever an rocking shaft linkage system. The strength and resiliency characteristics of the material used in the construction of the capsules is such that they expand or contract under the influence of varying external atmospheric pressure. The complete mechanism is housed within a case which, with exception of a static pressure connector at the rear, is completely sealed. A barometric pressure-setting device also forms part of the mechanism. Static pressure is exerted on the outside of the capsules and as changes in the pressure take place the capsules expand or contract; for example, as altitude increases, static pressure decreases and the capsules expand. The small deflections thus obtained are magnified through the lever and rocking shaft linkage and gear train system to produce an indication of the pressure changes in terms of aircraft altitude.

Since altimeters are calibrated to standard atmospheric conditions, any departure from the assumed values will change the pressure/ altitude relationship causing errors in indication. For example, if the prevailing atmospheric pressure at a sea-level aerodrome falls below the standard value, an altimeter situated at the aerodrome will respond correctly to the pressure change and indicate that the aerodrome now stands at a certain altitude above sea-level; in other words, the altimeter over-reads. Conversely, the instrument would under-read should the atmospheric pressure increase. In order to compensate for these barometric errors, a pressure setting device consisting primarily of an adjusting knob, a scale or digital counter and a gear mechanism, is interposed between the capsule stack and the pointers. The scale or counter may be calibrated in either millibars or inches of mercury, the readings being visible through an aperture in the main dial. When the adjusting knob is rotated, the scale or counter and the complete altimeter mechanism are also rotated, the gearing being so arranged that, for any atmospheric pressure setting, the altimeter pointer or pointers rotate to indicate the altitude equivalent of the pressure. Thus, to correct the reading of the altimeter in the example considered and so make the pointers indicate zero feet, the scale or counter must be set to the atmospheric pressure prevailing at the aerodrome.

The setting and correction of altimeters to known atmospheric pressure forms part of aircraft operating procedures and is of great importance for take-off, landing and for maintaining adequate height separation of aircraft and terrain clearance. Three Q Code for the transmission of meteorological and other operational information. They are:

a) Setting aerodrome atmospheric pressure so that an altimeter reads zero on landing and take-off (QFE).
b) Setting mean sea-level atmospheric pressure so that an altimeter reads the aerodrome altitude above mean sea-level (QNH).
c) Setting mean sea-level atmospheric pressure in accordance with ICAO standard atmosphere, i.e. 1013.25 millibars or 29.92 in HG (QNE).

Variations in atmospheric temperature affect the rate at which pressure changes with altitude relative to standard conditions; therefore, errors in the indications of altimeters can also occur due to such temperature variations.

Correction of these errors is applied in flight by means of a height calculator based on the slide-rule principal. The effects of temperature on pressure-sensing capsules and linkage mechanisms are compensated by devices operating on the bi-metallic principle.

The presentation of altitude by altimeters in current use varies from the multi-pointer type in which, the pointers are of different lengths and indicate ‘hundreds’, ‘thousands’, and ‘tens of thousand’ of feet against a common scale. In the drum/pointer type of presentation of pointer indicates ‘hundreds’ of feet against a fixed scale, while ‘thousands’ of feet are indicated by a drum which is proportionally rotated by the pointer mechanism. The drum scale is visible through an aperture in the instrument dial and is referenced against a datum marker across the aperture. In the digital counter/pointer method of presentation ‘thousands’ and ‘tens of thousands’ of feet are indicated by incremental changes of separate counters located across the centre of the dial. The counter/drum/pointer type of altimeter presentation provides a 5-digit numerical display of altitude with the last three digits forming the drum and showing ‘hundreds’ of feet. The pointer displays 1000 feet for each complete revolution with 20 feet scale markings. (fig.15.7).


*Fig. 15.7, Typical Altimeter Presentation.*

In order to avoid the mechanical loads which would otherwise be imposed on the pressure-sensing capsules by actuating several counters, a servo drive system may be used in altimeters employing these latter types of presentation, the movement of the capsules actuates the armature of an electromagnetic pick-off assembly producing a signal proportional to the prevailing pressure. This output signal is amplified by an external amplifier unit and supplied to a servomotor located within the indicator, to drive the pointer and counters through special gear assemblies. The amplified signal may, in some cases, also be used to drive synchros which in turn drive indicators. In the event of electrical power failure to the instrument, a power failure warning flag is incorporated to indicate that the altimeter is inoperative.

Because of the inherent pressure errors servo altimeters generally have a built-in correction system, 'tailored' for the particular aircraft design, that minimises this error for the full range of flight speeds and altitudes. Correction for datum pressure changes in the servo altimeter is achieved by mechanically displacing the datum of the electromagnetic pick-off assembly. The pick off senses this change and the system runs to null the datum change, thereby driving the pointer to zero.

The use of altimeters employing any of the three principal presentation methods, is governed primarily by the ease with which readings may be interpreted over the altitude ranges in which specific type of aircraft must operate. Multi-pointer presentations, particularly the three-pointer type, are very susceptible to misreading, e.g. 1000 ft for 10,000 ft, and various modifications have been incorporated to overcome this difficulty; for example, the type illustrated in Figure 15.7, employs distinctively shaped pointers and low altitude warning sectors. However, the use of multi-pointer altimeters as a standard for all types of aircraft is severely limited by performance characteristics and operational altitude ranges. For high-performance turbo-jet aircraft, in which misreading of altimeters may prove hazardous, the counter/pointer presentation is preferred.

**Encoding Altimeters**

To enable the altitudes of an aircraft to be known for air traffic control purposes, an airborne radar beacon transponder can be interrogated by a ground radar. These transponders have 4,096 codes available, so the encoding altimeters not only provide the flight crews with a visible read-out of the aircraft flight level but code the transponder so that it can reply to ground station with a signal providing a visible indication on a radar screen of the aircraft's altitude in 100 ft increments. Encoding altimeters of the non-servo type must have an extra low torque pick-off and the majority now in use employ optical encoders. In this system, the capsule drives a glass disc, etched with transparent and opaque sectors. A light source shines through the disc onto photoelectric cells which convert the disc's movement into coded signals for the transponder. This type of pick-off provides a high degree of accuracy with very low torque requirements.

**AIRSPEED INDICATORS**

These instruments are, in effect, sensitive pressure gauges which measure the difference between the pitot and static pressures, and present such differences in terms of indicated air speed. Indicators are made by various manufacturers and , although fulfilling identical roles, they may vary in their mechanical construction; however, the basic construction and operating principle is the same for all types.

The mechanism, which is contained within an airtight case, is comprised of a pressure sensitive capsule coupled to a pointer mechanism via a lever and rocking shaft linkage system. Two adapters, identified by the letters P and S, are located at the rear of the case and provide for the connection of the indicator to the pitot and static pressure pipelines respectively.

Pitot pressure is transmitted to the interior of the capsule via a short length of capillary tube connected between the capsule and pitot adapter, while the static pressure is transmitted directly to the interior of the indicator case and acts on the outside of the capsule. Changes in either of the two pressures establishes a differential across the capsule causing it to expand or contract. The small deflections of the capsule are transmitted through the lever an rocking shaft linkage system via a gear quadrant and pinion to the pointer mechanism, which produces a magnified angular deflection of the pointer or pointers over a scale calibrated in knots or in miles per hour.

**Maximum Allowable Air speed Indicators**

Another type of airspeed indicator in use is the maximum allowable air speed indicator which includes a maximum allowable needle to indicate a decrease in maximum allowable air speed with an increase of altitude. It operates from an extra capsule in the airspeed indicator which senses changes in altitude and measures this change on the dial of the instrument. Its purpose is to indicate maximum allowable indicated air speed at any altitude.

**True Airspeed Indicator**

The case of this instrument holds both an airspeed indicator which moves the pointer and an altimeter mechanism which moves the dial. The movement of the altimeter mechanism is opposed or aided by the action of a bimetallic spring exposed to outside air flow, and as the aircraft increases in altitude, the dial rotates in such a direction that the pointer will indicate a higher value. If the air is warmer than standard for the altitude at which the aircraft is flying, the temperature sensor will assist the altimeter to cause the true air speed reading to be higher than under standard temperature conditions.

**VERTICAL SPEED INDICATORS**

These indicators are sensitive differential pressure gauges which are connected to the static system and sense the rate of change static pressure and present it in terms of vertical speed (rate of climb and descent ) in feet per minute. The mechanism is housed in a cylindrical case which is rendered airtight except for the static pressure connector at the rear of the case. Basically the mechanism consists of a sensitive capsule, a 'calibrated leak' assembly or metering unit, a lever and rocking shaft linkage system, and a gear type pointer mechanism. The connection of the capsule and metering unit to the pressure connector is arranged so that static pressure is fed directly to the interior of the capsule and also allowed to 'leak' at a calibrated rate to the interior of the instrument case. Thus, when the static pressure varies due to changing attitude, the metering unit has restrictive effect causing the pressure change in the case to lag behind the pressure change in the capsule. The resulting differential pressure across the metering unit and the capsule causes the latter to expand or contract and drive the pointer via the linkage and gear mechanisms. The scale is calibrated to indicate climb and descent, in the clockwise and anti-clockwise directions respectively, from a zero graduation situated at the 9 o'clock position. A set screw, or in some instruments an adjusting knob, is provided in the lower left-hand corner of the bezel to permit the capsule datum position to be adjusted and thus reposition the pointer to zero via the linkage and gear mechanism within the limits set by the manufacturer.

In the level flight, the pressures inside the capsule and the instrument case remain the same; therefore the mechanism is at rest and the pointer indicates zero. When the aircraft climbs, the static pressure decreases inside the capsule but, due to the metering unit, the case pressure will remain the greater and so cause the capsule to contract and drive the pointer to indicate a rate of climb. The pressure difference thus established is maintained until any further alteration to rate of change of altitude takes place. During a descent the changes in static pressure are in the opposite sense causing the capsule to expand and to drive the pointer to indicate the appropriate rate of descent until level flight is resumed and no further altitude change takes place.

**INSTANTANEOUS VERTICAL SPEED INDICATOR**

The ordinary vertical speed indicator whose indication lags the pressure change would be of greater value if it had no lag, and for this reason the instantaneous vertical speed indicator has been developed. An instantaneous vertical speed indicator uses a vertical speed indicator mechanism in the case with an accelerometer-operated pump or dashpot across the capsule. When the aircraft noses over to begin a descent, the inertia of the accelerometer piston causes it to move upwards, instantaneously increasing the pressure inside the capsule and lowering the pressure at the metering unit. This gives an immediate indication of a descent. By the time the lag of the ordinary vertical speed indicator has been overcome so it will indicate, there is no more inertia from the nose-down rotation and the piston is again centred making the instruments ready to indicate instantly the levelling off from the descent.

**MACHMETERS**

In the operation of high performance aircraft capable of speeds approaching or exceeding that of sound, it is necessary to measure air speeds which are directly related to altitude and also to the variations in speed of sound which occur with atmospheric density variations. Therefore, in addition to conventional airspeed indicators, instruments integrating both speed and altitude measuring functions and referred to as Machmeters, are installed in aircraft of this type. Machmeters indicate the air speed in terms of Mach number which is defined as the ratio of the air speed of an aircraft to the speed of sound under the prevailing atmospheric conditions in which the aircraft is flying. The ratio may be expressed as a numerical percentage but for convenience is designated decimally and with the suffix M; thus, 0.8 M means a speed that is 80% of the speed of sound under the ambient conditions in which the aircraft is flying.

The mechanism of a Machmeter consists basically of an airspeed measuring unit an altitude measuring unit both of which are connected through calibration arms and sliding shafts to a common gear mechanism actuating a single pointer. The complete assembly is housed in a case provided with pitot and static pressure connectors at the rear. A lubber mark mounted over the instrument dial provides for the setting of the limiting Mach number specified for the aircraft in which the instrument is to be installed, and may be adjusted either by a screw or knob at the front of the instrument.

The capsule of the airspeed measuring unit expands and contracts in response to the difference between pitot and static pressure and deflects the pointer to positions related to corresponding air speeds. The altitude unit capsule expands and contracts in response to changes in static pressure only, and since it is interconnected with the airspeed measuring unit it determines the point of contact between the calibration arms thereby modifying the magnification ratio of the airspeed unit. Thus, the final deflected position of the pointer relates to a constant air speed at a particular altitude and, as atmospheric temperature variations are taken into account by the basic calibration formula, the Mach number as conventionally defined is therefore measured in terms of a pressure ratio.

■■■

# CHAPTER : 16
# FLIGHT INSTRUMENTS : GYROSCOPIC SYSTEMS

**SOURCES OF POWER FOR GYRO OPERATION**
The gyroscopic instruments can be operated either by a vacuum system or an electrical power source. In some aircraft, all the gyros are either vacuum or electrically motivated ; in others, vacuum (suction) systems provide the power for the attitude and heading indicators, while the electrical system drives the gyro for operation of the turn needle. Either alternating or direct current is used to power the gyroscopic instruments.

**Vacuum System**
The vacuum system spins the gyro by sucking a stream of air against the rotor vanes to turn the rotor at high speed, essentially as a water wheel or turbine operates. Air at atmospheric pressure drawn through a filter or filters drives the rotor vanes, and is sucked from the instrument case through a line to the vacuum source and vented to the atmosphere. Either a venturi or vacuum pump can be used to provide the vacuum required to spin the rotors of the gyro instruments.

The vacuum value required for instrument operation is usually between 3½ in. to 4 ½ in. Hg and is usually adjusted by a vacuum relief valve located in the supply line. The turn-and-bank indicators used in some installations require a lower vacuum setting. This is obtained using an additional regulating valve in the individual instrument supply line.

**Venturi-Tube Systems**
The advantages of the venturi as a suction source are its relatively low cost and simplicity of installation and operation. A light, single-engine aircraft can be equipped with a 2-in. venturi (2 in. Hg. vacuum capacity) to operate the turn needle. With an additional 8-in. venturi, power is available for the attitude and heading indicators. A venturi vacuum system is shown in figure 16.1.

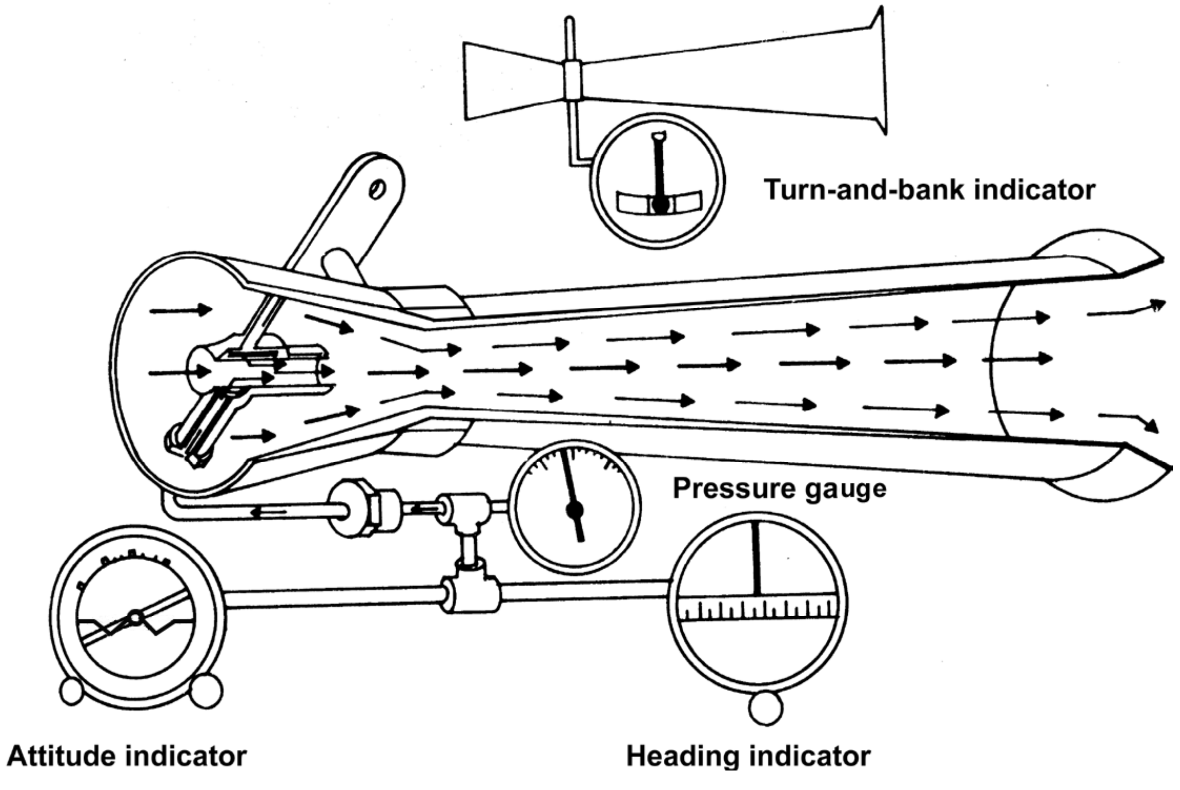


*Fig. 16.1, Venturi vacuum system.*

The line from the gyro (figure 16.1) is connected to the throat of the venturi mounted on the exterior of the aircraft fuselage. Throughout the normal operating airspeed range the velocity of the air through the venturi creates sufficient suction to spin the gyro.

The limitations of the venturi system should be evident from the illustration in figure 16.1 . The venturi is designed to produce the desired vacuum at approximately 100 m.p.h. under standard sea-level conditions. Wide variations in airspeed or air density, or restriction to airflow by ice accretion, will affect the pressure at the venturi throat and thus the vacuum driving the gyro rotor. And, since the rotor does not reach normal operating speed until after takeoff, pre flight operational checks of venturi-powered gyro instruments cannot be made. For this reason the system is adequate only for light-aircraft instrument training and limited flying under instrument weather conditions. Aircraft flown throughout a wider range of speed, altitude, and weather conditions require a more effective source of power independent of airspeed and less susceptible to adverse atmospheric conditions.

**Engine-Driven Vacuum Pump**

The vane-type engine-driven pump is the most common source of vacuum for gyros installed in general aviation light aircraft. One type of engine driven pump is mounted on the accessory drive shaft of the engine, and is connected to the engine lubrication system to seal, cool, and lubricate the pump.

Another commonly used source of vacuum is the dry vacuum pump, also engine-driven. The pump operates without lubrication., and the installation requires no lines to the engine oil supply, and no air-oil separator or gate check valve. In other respects, the dry pump system and oil lubricated system are the same.

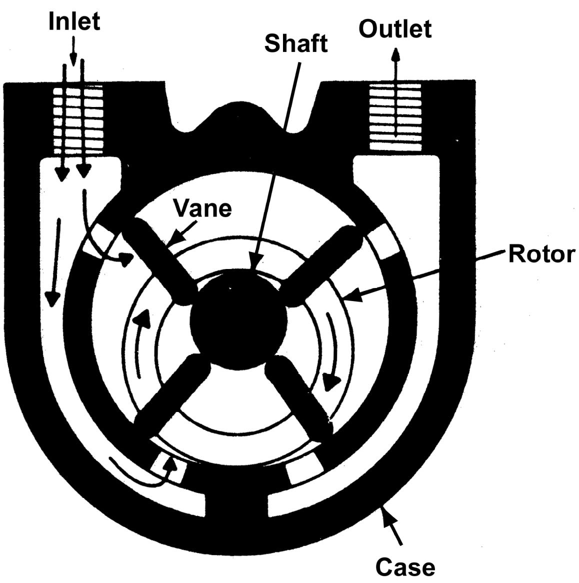



*Fig. 16.2, Cutaway view of a vane-type engine driven vacuum pump.*

The principal disadvantage of the pump-driven vacuum system relates to erratic operation in high-altitude flying. Apart from routine maintenance of the filters and plumbing, which are absent in the electric gyro, the engine-driven pump is as effective a source of power for light aircraft as the electrical system.

**Typical Pump-Driven Vacuum System**

Figure 16.3 shows the components of a vacuum system with a pump capacity of approximately 10" Hg at engine speeds above 1000 rpm. Pump capacity and pump size vary in different aircraft, depending on the number of gyros to be operated.

***Air-Oil Separator***

Oil and air in the vacuum pump are exhausted through the separator, which separates the oil from the air; the air is vented outboard, and the oil is returned to the engine sump.

***Suction Relief Valve***

Since the system capacity is more than is needed for operation of the instruments, the adjustable suction relief valve is set for the vacuum desired for the instruments. Excess suction in the instrument lines is reduced when the spring-loaded valve opens to atmospheric pressure. (See fig. 16.4)

***Pressure Relief Valve***

Since a Reverse flow of air from the pump would close both the gate check valve and the suction relief valve, the resulting pressure could rupture the lines. The pressure relief valve vents positive pressure into the atmosphere.

***Gate Check Valve***

The gate check valve prevents possible damage to the instruments by engine backfire, which would reverse the flow of air and oil from the pump. (See fig. 16.5)

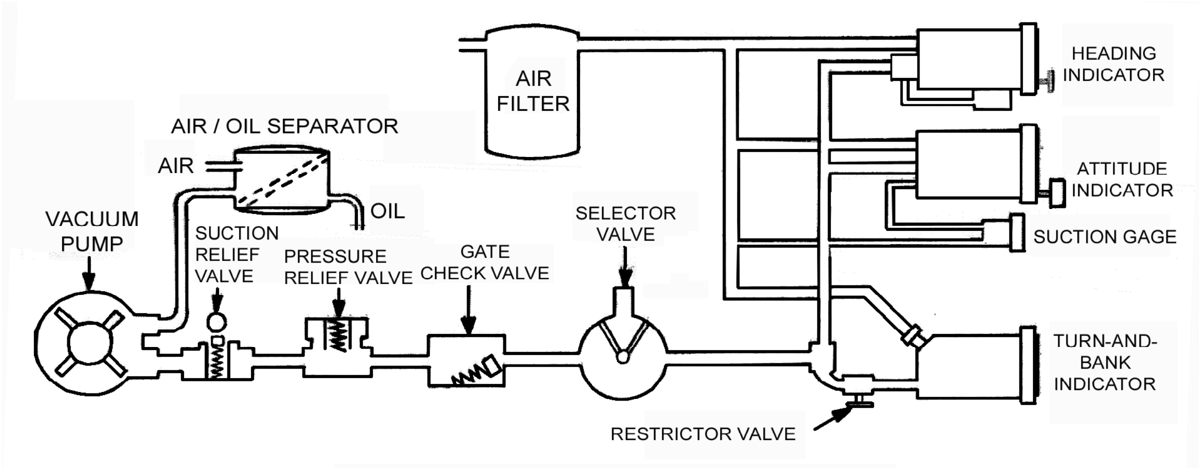



*Fig. 16.3, Typical pump-driven vacuum system.*

### *Selector Valve*

In twin-engine aircraft having vacuum pumps driven by both engines, the alternate pump can be selected to provide vacuum in the event of either engine or pump failure, with a check valve incorporated to seal off the failed pump.

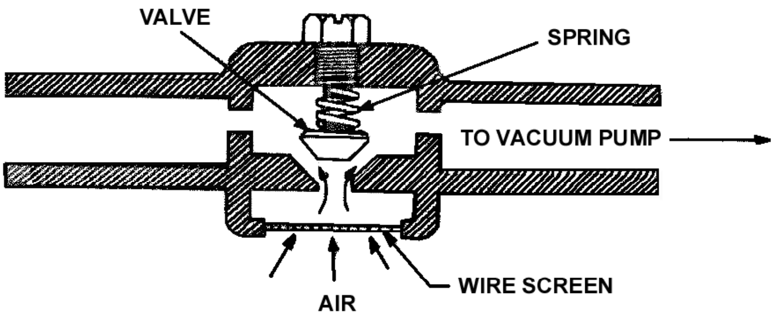



*Fig. 16.4, Vacuum regulator valve.*

### *Restrictor Valve*

Since the turn needle operates on less vacuum than that required for other gyro instruments, the vacuum in the main line must be reduced. This valve is either a needle valve adjusted to reduce the vacuum from the main line by approximately one-half, or a spring-loaded regulating valve that maintains a constant vacuum for the turn indicator, unless the main line vacuum falls below a minimum value.

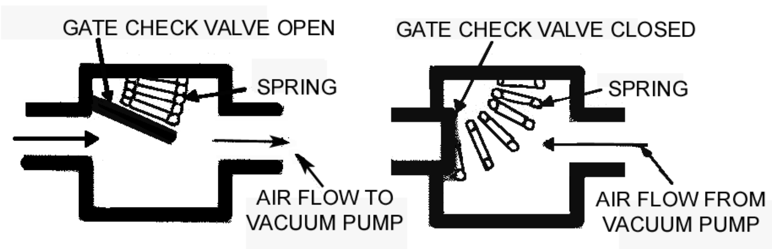



*Fig. 16.5, Gate check valve.*

### *Air Filter*

The master air filter screens foreign matter from the air flowing through all the gyro instruments, which are also provided with individual filters. Clogging of the master filter will reduced airflow and cause a lower reading on the suction gauge. In aircraft having no master filter installed, each instrument has its own filter. With an individual filter system, clogging of a filter will not necessarily show on the suction gauge.

### *Suction Gauge*

The suction gauge is a pressure gauge, indicating the difference in inches of mercury, between the pressure inside the system and atmospheric or cockpit pressure. The desired vacuum, and the minimum and maximum limits, vary with gyro design. If the desired vacuum for the attitude and heading indicators is 5" and the minimum is 4.6", a reading below the latter value indicates that the airflow is not spinning the gyros fast enough for reliable operation. In many aircraft, the system provides a suction gauge selector valve, permitting the pilot to check the vacuum at several points in the system.

**Suction**

Suction pressures discussed in conjunction with the operation of vacuum systems are actually minus or negative pressures (below sea level). For example, if sea level equals 16.5 p.s.i. then 1" Hg (1 inch mercury) or 1 p.s.i. vacuum is equal to -1 p.s.i. negative pressure or 16.5 positive pressure. Likewise 3" Hg = -3 p.s.i. negative pressure or +14.5 positive pressure.

Of course, for every action there is an equal and opposite reaction. Therefore when the vacuum pump develops a vacuum (negative pressure) it must also create pressure (positive). This pressure (compressed air) is sometimes utilized to operate pressure instruments, deicer boots and inflatable seals.

**Typical System Operation**

The schematic of a Vacuum system for a twin-engine aircraft is shown in figure 16.6. This vacuum system consists of the following components ; two engine-driven pumps, two vacuum relief valves, two flapper-type check valves, a vacuum manifold, a vacuum restrictor for each turn-and-bank indicator, an engine four-way selector valve, one vacuum gauge, and a turn-and-bank selector valve.

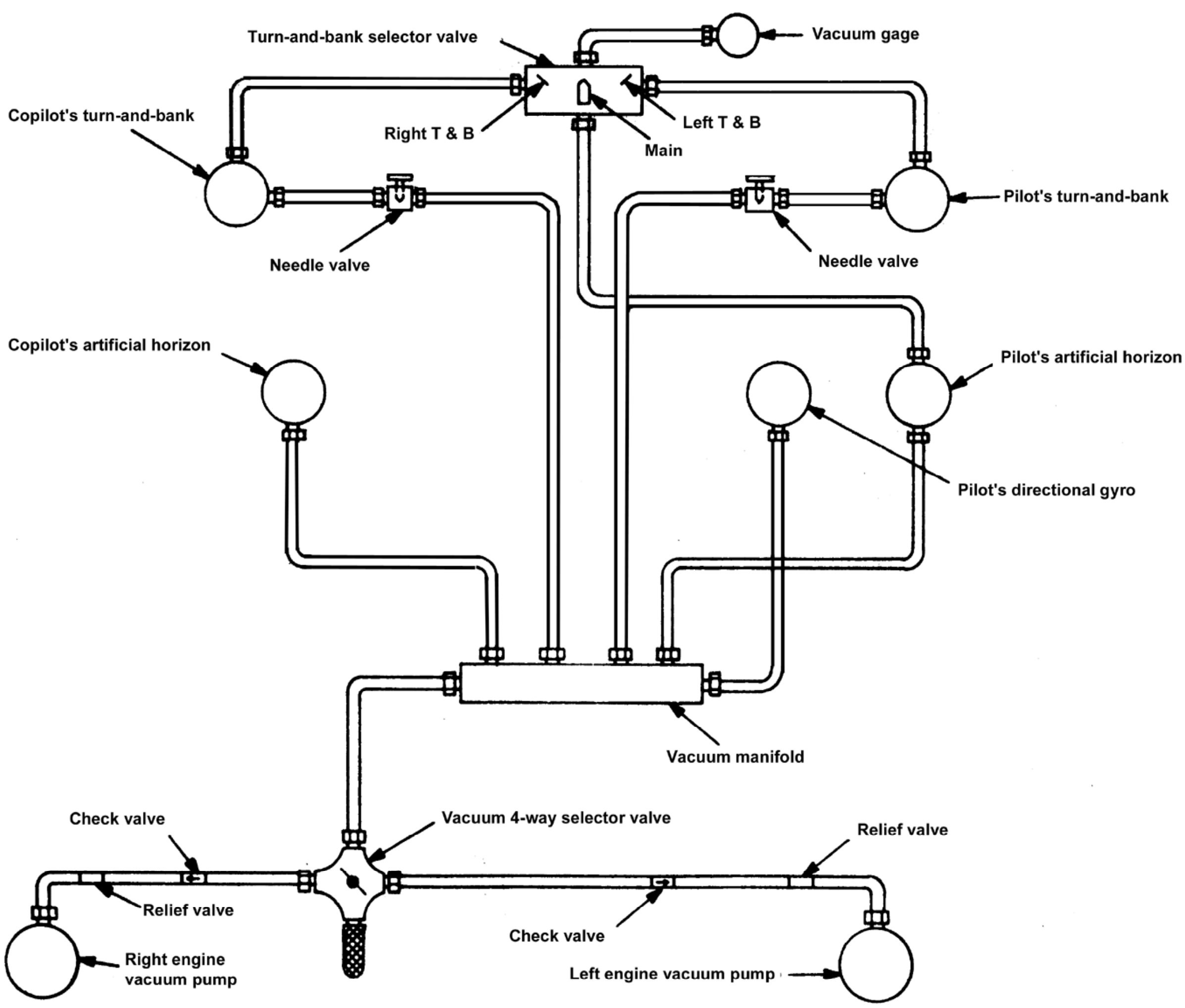



*Fig. 16.6, Vacuum System for a multi-engine aircraft.*

The left and right engine-driven vacuum pumps and their associate lines and components are isolated from each other, and act as two independent vacuum systems. The vacuum lines are routed form each vacuum pump through a vacuum relief valve and through a check valve to the vacuum four-way selector valve.

From the engine four-way selector valve, which permits operation of the left or right engine vacuum system, the lines are routed to a vacuum manifold. From the manifold, flexible hose connects the vacuum-operated instruments into the system. From the instrument, lines routed to the vacuum gauge pass through a turn-and-bank selector valve. This valve has three positions : main, left T & B, and right T & B. In the main position the vacuum gauge indicates the vacuum in the lines of the artificial horizon and directional gyros. In the other positions, the lower value of vacuum for the turn-and-bank indicators can be read.

**GENERAL**

Information on the attitude and direction of an aircraft is provided by the gyro horizon, turn and slip indicator and direction indicator. These instruments depend for their operation on the fundamental properties of a gyroscope designed to establish stabilised references which, in conjunction with the indications of pilot-static instruments are required for various condition of flight operations.

For those not familiar with the basic principles of these flight instruments, brief operating details are given in paragraphs dealing with the individual instruments.

**GYROSCOPES**

A gyroscope (normally abbreviated to gyro) is a rotating mass having freedom in one or more planes at right angles to the plane of rotation. It possesses two fundamental characteristics: gyroscopic inertia or rigidity, and precession. Gyroscopic Inertia is the property a rotating mass has of reluctance to change its plane of rotation in space unless acted upon by an external force (Newton's first law of motion; every body continues in its state of rest or of uniform motion in the straight line unless it is compelled by forces to change that state). Precession is the angular change of direction of the plane of rotation, under the action of an external force.

**DEFINITIONS**

**Free Gyro**

A gyro having complete freedom in three planes at right angles to each other. This is also sometimes known as a 'space' gyro.

**Tied Gyro**

A gyro having freedom in three planes at right angles to each other but controlled by some external source.

**Earth Gyro**

A tied gyro controlled by gravity to maintain its positions relative to the earth.

**Rate Gyro**

A gyro having one plane of freedom at right angles to the plane of rotations, so constructed as to measure rate of movement about the plane at right angles to both the plane of rotation and the plane of freedom.

**PRACTICAL APPLICATIONS**

Mechanically, a wheel can only be mounted so that it has complete freedom in three planes by mounting it in a system of rings or gimbals. The spinning rotor in aircraft instruments constitutes the gyro. All practical applications of the gyro are based on the two characteristics: gyroscopic inertia, and precession.

**Gyroscopic Inertia**

When the rotor of a free gyro is spinning it will maintain its plane of rotation fixed in space and will not be affected by movement of its outer gimbal system.

**Precession**

When a torque is applied to disturb the plane of rotation of a gyro, the gyro will resist angular movement in the plane of that torque, but will move in the plane at right angles to the disturbing torque; this resulting movement is called precession. The direction of the precessional movement is dependent on the direction of the disturbing force and the direction of rotation of the gyro, and can be determined by considering the applied torque as due to a force acting on a point on the rim of the gyro rotor at right angles to the plane of rotation. If that imaginary point is carried around the rotor 90° in the direction of rotation, that will be the point at which the force is apparently taking effect, moving that part of the rotor rim in the same direction as the disturbing force. (fig.16.7).

When the gyro rotor is spinning, it maintains its plane of rotation fixed in space, but there is no definite attitude in space which the gyro will naturally assume. Therefore, if the gyro is to be used as a reference it must be controlled, so that the gyro assumes a definite attitude either in space or relative to the earth, after which it will be stabilised by its own inertia. A gyro fitted with such a device is called a 'tied' gyro. If tied by gravity control, it is called an 'earth' gyro. An example of a tied gyro is the directional gyro which has its axis of rotation maintained in a plane at right angles to the outer gimbal ring. An example of an earth gyro is the gyro horizon which has its axis of rotation maintained in a vertical position relative to the earth's surface, by pendulum control.

**Rate Gyros**

If a gyro is mounted so that it has freedom about two axes only it can be adapted to be used as a 'rate' gyro, i.e. to measure rate angular movement. If there is an angular movement in the plane in which the gyro has no freedom, the rotor will be precessed until its plane of rotation coincides with the plane of angular freedom and its direction of rotation coincides with the direction of the angular movement.

In order to obtain an indication of rate, it is necessary to restrict precession of the rotor so as to apply a torque to it which will cause precession in the same plane as the angular movements. This is usually done by springs. On turning the base, the rotor will precess until the spring applies a force to the rotor sufficient to cause it to precess about the turning axis at exactly the same rate as the base is being turned. The amount of precession against the spring will be dependent on the rate of angular movement (fig.16.8).

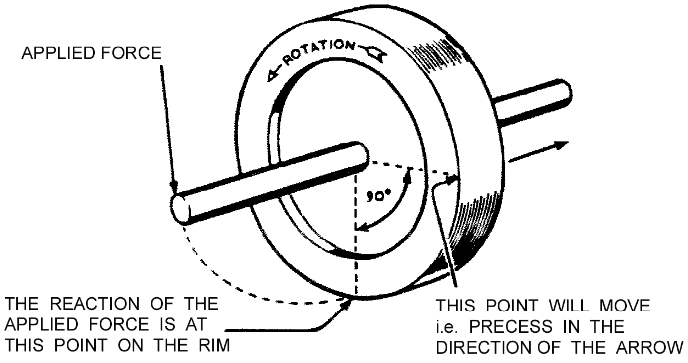



*Fig. 16.7, Precession.*

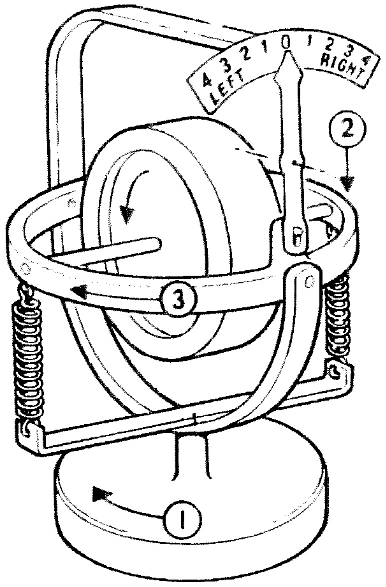



*Fig. 16.8, A basic rate gyro.*

Turning the base of the gyro illustrated in Figure 16.8 above in the direction of arrow 1 will cause the rotor to precess so that the needle will move over to the right-hand side of the scale. The right-hand spring will now exert a force down (as shown by arrow 2) This force will process the rotor to turn in the same plane and at the same rate as the base. Since the base and rotor axis are fixed in relation to each other, any change in the rate of angular movement of the base must cause a change in the precessional rate of the rotor. This rate of angular movement of the base must always be balanced by precession of the rotor in the same plane. This balance is achieved by precession against the spring to produce the requisite force, and any given rate of turn will be shown as a definite amount of movement of the needle over the scale provided the rotor speed is constant.

## Gyro Horizons

Gyro horizons utilise vertical spin axis gyros and provide visual indications of any change of pitch and roll attitude of an aircraft by the attitude of fixed reference relative to gyro-stabilised references. Pitch attitude is indicated by the position of a symbol, representing the aircraft, relative to a stabilised bar symbolising the natural horizon. The bar is pivoted on the outer gimbal ring of the gyro in such a way that relative positions are magnified. Simultaneously the bank attitude of the aircraft is indicated by the positions of a fixed bank angle scale relative to a bank pointer which is also gyro-stabilised. The gyro assembly is pivoted about the longitudinal axis of the instrument case to provide freedom of movement about all three axes of the aircraft.

With the instrument in operation, gyroscopic rigidity maintains the spin axis in its normal operating position and the horizon bar with respect to the aircraft symbol. When the bank attitude of the aircraft changes the instrument case and aircraft symbol move together with the aircraft about the longitudinal axis, and will lie at an angle to the horizon bar. At the same time the bank angle scale moves with the aircraft, indicating against the bank pointer the number of degrees of bank.

If the aircraft changes its attitude in pitch, the case and aircraft symbol move together with the aircraft about the lateral axis of the gyro so that the aircraft symbol will be above or below the horizon bar to respectively indicate a climb or descent.

The gyro assemblies of all gyro horizons incorporate a levelling device which maintains the gyro in its normal operating position. In some electrically-operated versions additional fast erection systems are incorporated which are controlled by a knob located at the front of the instrument. The operation of a fast erection system varies but, in general it may be based on mechanical locking or caging of the gimbal system, or on the principles of increasing the current applied to the normal levelling device.

**TURN AND SLIP INDICATORS**

These instruments indicate the rate of turn of an aircraft and also any slip or skid during a turn. Two separate indicating mechanisms are utilised; a turn indicating mechanism consisting of a gyro-controlled pointer referenced against a turn scale, and gravity-controlled slip indicating mechanism which may be of the ball-in tube inclinometer type, or of the gravity weight and pointer type. The gyro is of the two-axis type, known as a rate gyro, and consists of a single gimbal ring supporting the rotor about a horizontal axis. The gimbal ring is pivoted about the longitudinal axis of the instrument case. A calibrated rate spring is connected between the gimbal ring and the case. The pointer may be coupled directly to the gimbal ring, or coupled to it by means of a gear mechanism, depending on the manufacturer's design. Transient movements of both turn and slip indicating mechanisms are minimised by special damping devices.

When the aircraft turns, a force is exerted on the gyro which precesses the gyro about its longitudinal axis against the tension of the rate spring until equilibrium is established. The displacement of the gyro is therefore related to the rates of performance and aircraft turn, the latter being indicated on the scale by the deflection of the pointer on the appropriate side of the scale's zero positions.

The slip indicating mechanisms depend upon the effects of gravitational and centrifugal forces for their operation. During straight flight in the normal lateral attitude gravity acts vertically through the mechanism, and the ball or pointer remains at the central datum. In turning flight, centrifugal forces are added vectorially to those of gravity and the indicating mechanism will therefore indicate the resultant or direction of, apparent gravity, when turns are correctly banked the forces are such that the apparent gravity will maintain the ball or pointer at central datum. If a slip or skid occurs the apparent gravity will be deviated causing a displacement of the ball or pointer in the direction of slip or skid by a related amount.

**DIRECTION INDICATORS**

Direction indicators provide a stabilised directional reference for maintaining a desired course and for turning on to a new heading. They are also used as complementary instruments to direct-reading magnetic compasses. The instrument is nonmagnetic and consists of a gyro pivoted about the vertical axis of the case so that the rotor spins about a horizontal axis. A rectangular opening in the front of the case carries a fixed lubber line which is referenced against a circular card graduated in degrees and secured to the outer gimbal ring of the gyro.

With the instrument is in operation, gyroscopic rigidity stabilises the gyro and circular card and a certain heading is registered against the fixed lubber line. If the aircraft heading changes, the instrument case and lubber line turn with the aircraft about the stationary gyro and card, thus giving an indication of the number of degrees through which the aircraft turns. Headings corresponding to those indicated by the magnetic compass are set in direction indicators by means of a setting knob at the front of the instrument. When the knob is pushed in, it engages a locking mechanism with the gyro assembly which can then be rotated to the desired heading by turning the setting knob. The knob has a secondary function of caging the gyro assembly to prevent damage during transit, installation and removal.

■■■

# CHAPTER : 17
# DIRECT-READING MAGNETIC COMPASSES

**COMPASS ERRORS AND METHODS OF COMPENSATION**
In connection with the compensation of direct-reading compasses, the following principal errors have to be taken into account:-

**Index Error**
This error, which is also known as Coefficient “A” error, results from malalignment of a compass in its mounting and has the same magnitude on all headings. The error is calculated by averaging the algebraic sum of deviations on each of cardinal and quadrantal headings. Compensation is effected by rotating the compass in its mounting through the number of degrees calculated, and relative to the longitudinal axis of the aircraft.

**One-Cycle Errors**
These refer to the deviations produced in compass readings as a result of the effects of components of permanent or hard-iron magnetism of the aircraft’s structure. The deviations vary as sine or cosine functions of the aircraft’s heading, the maximum deviations being termed Coefficients “B” and “C” respectively. Other sources of these errors are components of soft-iron magnetism induced by the earth’s field, the hard iron itself and the effect of electric currents from cables or equipment which may be mounted in the vicinity of compass. The Coefficient “B” error is calculated by averaging the algebraic difference of the deviations on the East and west headings, while the Coefficient “C” error is calculated by averaging the algebraic difference of deviations on the north and south heading. Compensation is effected by a permanent magnet type compensator unit, which depending on the type of compass, is either secured direct to the compass bowl or is mounted separately on the compass support bracket. The unit contains two pairs of magnets, the axes of which are so disposed that their fields neutralize the effects of the magnetic components producing “B” and “C” deviations. Each pair of magnets can be rotated by a shaft provided with either a screwdriver slot or a shaped end requiring the use of a key.

**TYPES OF COMPASS**
Two types of direct-reading magnetic compass are in use, the card type and the grid-steering type. The major differences between the two are in the magnet system arrangement, the method of heading presentation, and the arrangement of the deviation compensator devices.

Card type compasses, which are designed for mounting on an instrument panel or on a coaming panel, indicate magnetic headings by means of a graduated card affixed to the magnet system and registering against a lubber line in the front of the bowl. The deviation compensator device is usually secured directly to the compass bowl.

Grid steering type compasses employ a needle and filament type magnet system which is referenced against a grid-ring located over the compass bowl. The grid-ring, which may be rotated and clamped in any position, has a graduated scale and two pairs of parallel grid wires in the form of an open T. Magnetic headings are indicated by the number of degrees read against a lubber line in the compass bowl, when the needle and filaments lie parallel to the grid wires, and the North-seeking filament points to the North mark on the grid ring. These compasses may be designed for mounting on a bracket below and instrument panel or for inverted mounting in a cockpit roof. In the latter case, heading indications are observed by means of a mirror attachment. Deviation compensator devices are normally separate and are mounted on the compass supporting bracket.

**LOCATION**
The location of a compass in an aircraft is important and factors such as angle of observation, illumination, vibration, and in particular, the effect of magnetic disturbances require carefull consideration. The location is determined during the aircraft’s design stage and should not be altered unless authorised by the Airworthiness Authority.

Compensation may be made for reasonable amount of permanent magnetism, but variable sources of deviation must be kept distant in order to minimize their effects.

Where practicable, magnetic steel parts, especially movable parts, should not be positioned near the compass.

Electrical cables carrying unidirectional current produce a magnetic effect on the compass magnet system which is governed by the current and distance from the compass. All such cables should be positioned, if possible, at least 2 feet away from the compass. If double pole cables are used (i.e. supply and return cables run closely together) the magnetic effects of the cables are usually insignificant.

To minimize the effect on the compass of items of equipment with magnetic fields, such items should not be located closer to the compass than the relevant compass safe distance specified for each item by the manufacturer.

The possibility of magnetic fields generated by electric currents passing through windscreen pillars and frames, or through instrument mountings, should not be overlooked.

The effect of modifications to instrument installations, radio installations, electrical control panels or wiring in the vicinity of the compass must be considered, and tests should be made to determine whether any deviation will be caused under operating conditions.

In some instances, particularly in light aircraft, certain components and parts of the structure, e.g. control columns, control arches, tubular frames, may exhibit residual magnetism in varying amounts which cannot be corrected by compass deviation compensating devices. The origin and cause of such magnetism must be investigated and the appropriate remedial action must be taken. A simple practical detection method is to position a small pocket compass near suspect components or parts and to note any deflections of the needle.

**COMPASS SWINGING AREA**

Since the compass swinging procedure determines deviations caused by magnetic fields of an aircraft, it is necessary for the swinging to be undertaken at a location where only these fields and the earth's magnetic field can affect the compass readings. The location must, therefore, be carefully chosen and surveyed to prove that it is free form any interfering local magnetic fields.

**PREPARATION BEFORE SWINGING**

A check should be made to see that all airborne equipment is installed in the aircraft. Loose items or tools made from magnetic materials should not be left in the aircraft or carried by the personnel engaged in the swinging procedure. Any detachable cockpit mechanical control locks which might be magnetic should be removed and placed in their flight stowages. Where towing arms and towing vehicles are to be used for manoeuvring the aircraft, their possible magnetic effect should be investigated, and if significant they should be disconnected and moved clear before taking any compass readings.

Ensure that all equipment required for the swing is available, e.g. appropriate datum compass, sighting equipment and non-magnetic tools needed for deviation compensations.

Where appropriate, landing gear ground locks should be in position and landing gear shock struts should be checked to ensure that they are properly inflated. In some types of aircraft, a landing gear level latch is employed which is solenoid-operated. The solenoid is normally energized in flight, and since its magnetic filed may have on effect on the accuracy of the standby direct-reading compass, it should be also energized during the swinging procedure.

The flying controls should be in the normal straight and level flight positions when taking the readings, and should then be operated to ascertain that the movements have no adverse effect on the compass readings. Flaps, throttles, etc., should also be set to their "in flight" positions.

Electrical equipment, e.g. radio, instruments and Pitot tube heaters, should be switched on to ascertain that there are no adverse effects on the compass, In this connection, reference should always be made to the relevant aircraft manuals for details of the electrical loads to be selected appropriate to the aircraft operating conditions, e.g. normal day or night operation, or operation with emergency power.

Deviation compensator devices should be set to their magnetic neutral positions after installation, or after replacement of a compass or deviation compensator device where this is a separate unit.

**COMPASS SWINGING PROCEDURE**

Where compasses are to be compensated on a base with marked headings, the longitudinal axis of the aircraft must be aligned either on, or parallel to, the markings, usually with the aid to plumb lines dropped from points fore and aft along the axis. A datum compass such as the medium landing compass may also be used, whether or nor a concrete or Tarmac base is available. The datum compass should be aligned with the aircraft's longitudinal axis and positioned on the datum circle, or if this is not marked, at a specified distance from the aircraft (typical distances are 50 to 150 feet). In order that the longitudinal axis of the aircraft may be accurately determined, datum points on the aircraft and directions from which they are to be sighted, must be carefully selected. Some aircraft have provision for the fitment of sighting rods to aid determination of the longitudinal axis. The most straightforward direction for sighting is from the rear; the heading then corresponds to the datum compass reading. If the view from this aspect is unsatisfactory, the view from ahead of the aircraft should be considered, bearing in mind that reciprocal headings will be indicated. It is not advisable to take sights from positions at angles to the longitudinal axis.

Compasses installed in aircraft with fuselage-mounted engines should be compensated with engines running. If this is not practical then they should at least be re-checked on four equally spaced headings with the engines running, on completion of the swinging procedure.

■■■

# CHAPTER : 18
# REMOTE-READING COMPASSES

## PRINCIPLE OF REMOTE - READING COMPASS SYSTEMS

The principal component of any system is a flux detector unit, sometimes called a flux valve or flux gate. It is located in an area relatively free from any disturbing magnetic fields of the aircraft itself (e.g. a wing tip or vertical stabilizer) so that the horizontal component of the earth's magnetic field can be more accurately detected by the sensing element within the unit. The sensing element forms part of a synchro type of transmission system which, in most compasses, is coupled to the horizontal-axis directional gyro contained within either a heading display indicator mounted on the main instrument panel, or a master gyro unit from which heading data is transmitted to a separate indicator. In some aircraft using an inertial navigation system, the flux detector sensing element is connected to a compass coupler unit instead of a directional gyro, the purpose of the unit being to develop a stabilized magnetic heading reference from both sensing element and inertial navigation system signals. The sensing element is pendulously suspended in such a way that it has a limited amount of freedom in the pitching and rolling planes, but has no freedom in the yawing plane. In one currently used system the element is stabilized by a vertical gyro.

The sensing element is made up of material having high magnetic permeability wound with an exciter or primary coil and three pick-off or secondary coils. The exciter coil is supplied with a low-voltage single-phase a.c. at a constant frequency (typical values are 26 V, 400 Hz) and produces an alternating flux in the sensing element material. In addition to this flux, this horizontal component of the earth's magnetic field is also introduced; its effect being to change the total flux cutting the pick-off coils in such a manner that an e.m.f. is induced in them which, in terms of amplitude and phase, represents the magnetic heading.

The induced e.m.f. causes current to flow to the stator windings of receiver synchro within either the display indicator, master gyro unit or compass coupler, as appropriate, and a field is set up across the stator in a direction determined by the current flow in the windings. If the detector sensing element and receiver synchro are in synchronism, the synchro rotor is in its 'null' position and so no signal voltage is induced in its winding by the stator field cutting it. When a change in aircraft heading takes place, however, the position of the detector sensing element with respect to the earth's field also changes with the result that the current flow in the receiver synchro stator changes, causing the stator field to rotate. This, in effect, is the same as a rotor displacement from the 'null' position and although the rotor itself always tends to rotate with the stator field it is restrained momentarily by the mechanical coupling between it and the gyro. Thus, an error voltage is induced in the rotor winding; the phase and amplitude of which being dependent on the direction and magnitude of displacement of the rotor from the 'null' position. The voltage is fed to an amplifier and finally to a slaving system which produces a torque to precess the gyro and its indicating element to indicate the heading change. At the same time, the synchro rotor rotates in synchronism with the stator field.

## COMPASS ERRORS AND METHODS OF COMPENSATION

In connection with the swinging and compensation of remote-reading compasses, the following principal errors have to be taken into account:-

### Index Error

This error, which is also known as Coefficient "A" error, results from malalignment of the flux detector unit and has the same magnitude on all headings. The error is calculated by averaging the algebraic sum of the deviations on each of the cardinal and quadrantal headings.

### One-Cycle Errors

These refer to the deviations produced in compass readings as a result of the effects of components of permanent or hard-iron magnetism of the aircraft's structure. The deviations vary as sine or cosine functions of the aircraft's heading the maximum deviations being termed Coefficients "B" and "C" respectively. Other sources of these errors are, components of soft-iron magnetism induced by the earth's field, the hard iron itself, and electric currents from cables or equipment which may be mounted in the vicinity of the flux detector unit. The error due to Coefficient "B" is calculated by averaging the algebraic difference of the deviations on the East and West headings, while the Coefficient "C" error is calculated by averaging the algebraic difference of deviations on the North and South headings.

### Two-Cycle Errors

These errors result from imperfections in the transmission of heading data and are usually referred to as transmission errors. They can be cased by impedance or voltage imbalance in the flux detector sensing element or in the synchros of the compass system. Another source of two-Cycle error is soft-iron magnetism.

### Crosstalk Errors

These errors occur particularly during an "electrical" swinging procedure when the d.c. signals simulating the earth's field are applied. They are caused by different sensitivities of flux detector unit coils and by unequal air gaps separating the flux collector horns; the overall effect being to produce quadrature components which offset the field from the originally intended.

**Methods of Compensation**

Typical methods of compensating errors are descried in the following paragraphs:

**Index Errors Compensation**

The error to be compensated is calculated from the Coefficient “A" formula. In the most commonly used method, compensation is effected by rotating the flux detector unit in its mounting through the number of degrees so calculated, relative to a datum parallel to the longitudinal axis of the aircraft. The flux detector unit is rotated clockwise (when viewed from above) for a +A error and anticlockwise for a -A error. In some compass systems the flux detector is first aligned with the centre line of the aircraft, and the “A”’ error is removed by providing an electrical differential by means of a differential synchro between the flux detector unit and its synchro.

**One -Cycle Error Compensation**

The errors to be compensated are calculated from the Coefficient “B” and Coefficient “C” formulae. Depending on the type of compass system installed, compensation may be effected by one of the methods described in (a) and (b).

(a) ***Mechanical Methods***

In several types of compass systems, a permanent magnet type compensator unit is secured to the top of the flux detector unit. The compensator contains two pairs of magnets, the axes of which are so disposed that their fields neutralize the effects of the magnetic components producing “B” and “C’ deviations. Each pair of magnets can be rotated by a shaft containing a screwdriver slot, and associated gearing. In one particular type of compass system, mechanical compensation is effected by screws, a metal cam ad cam follower; the complete device being incorporated within the compass indicator. For details of this method of compensation reference should be made to the relevant Maintenance Manuals.

(b) ***Electrical Methods***

Electrical compensation is normally effected by a compensator unit connected to, and located remote from, the flux detector unit. A compensator unit contains two adjustable potentiometers, one for each coefficient. Depending on the compass system installed, the potentionmeters can be adjusted to vary either the electromagnetic fields produced in two coils mounted on top of the detector unit, or they can be adjusted to produce the correcting electro magnetic field within the flux detector pick-off coils, by supplying direct current to the coils themselves. In this latter case, the coils have the dual and simultaneous function of picking-off voltage resulting from heading changes and of deviation compensation. In certain types of compensator unit, test points are provided to permit measurement of the d.c. voltages across the two potentiometers.

**Crosstalk Error Compensation**

Compensation is effected during an “electrical” swing procedure, by applying d.c. signals to the flux detector unit coils and generating fields which oppose the quadrature components in the North-South and East-West directions.

**COMPASS SWINGING PROCEDURES**

The procedure to be adopted depends primarily on the type of compass and the method by which magnetic heading reference datum is obtained, i.e. form a base having headings marked out on it, form a datum compass, or form a base established for carrying out “electrical” swinging as outlined. Details of the procedure appropriate to a specific system and aircraft type are given in the Maintenance Manuals, and reference must, therefore, always be made to such documents. The information given in the following paragraphs is intended to serve only as a general guide to procedures and to certain associated important aspects.

**Compass Swinging Area**

Since the swinging procedure determines deviations caused by the magnetic fields of an aircraft, it is necessary for it to be undertaken at a location where only these fields and the earth’s magnetic field can affect the compass readings. The location must, therefore, be carefully chosen and surveyed to prove that it is free form any interfering local magnetic fields.

**Aircraft Sighting Points**

In all swinging procedures it is necessary to determine the position of the longitudinal axis of the aircraft with respect to a magnetic heading reference datum, and for this reason, two datum points on the aircraft (e.g. the aircraft nose and tip of the vertical stabilizer) and directions from which they are to be sighted, must be carefully selected. If the datum points are at different heights above the ground, and if the aircraft’s is rolled out of a level plane as a result of the compass base not being level, then an error can occur in the measured datum heading. Roll angles are normally small and the error approximates to :-

$$\text{Heading Error...} \quad \frac{\text{degree}}{\text{degree of roll}} \quad ...... \quad \text{is} \quad \frac{\text{Vertical distance between two points}}{\text{Horizontal distance between two points.}}$$

In several types of large transport aircraft, sighting is facilitated by the provision of sighting devices which are attached to the aircraft prior to carrying out the appropriate swinging procedure. Some examples are described in the following paragraphs.

**Sighting Rods**

In this example, front and rear sighting rods are attached to corresponding points provided along the centre line at the underside of the fuselage. An extension rod is also provided for attachment to either the rear sighting rod, if datum compass sighting are to be taken from the front of the aircraft, or the front sighting rod, if sighting are to be taken from the rear.

**Telescopic Target Fixture**

a) A typical fixture is shown in Fig.18.1. It is attached to the bulkhead of a main landing gear wheel well so that the scale and crosshair are on the inside of the turning circle of the aircraft. The setting of the fixture with respect to the longitudinal axis of the aircraft, is determined by sighting the telescope on a target mark located at the underside of the front fuselage, in this case at the forward jack pad position. The scale is of the centre-zero type, angles to the right being positive and angles to the left being negative when sighted from the datum compass. Magnetic heading of the aircraft is determined during the swinging procedure by sighting the datum compass. On the crosshair and scale of the target fixture, and calculating the difference between the scale reading and the datum compass reading, the latter being the bearing of the pre-selected reference target located at some distance from the base.

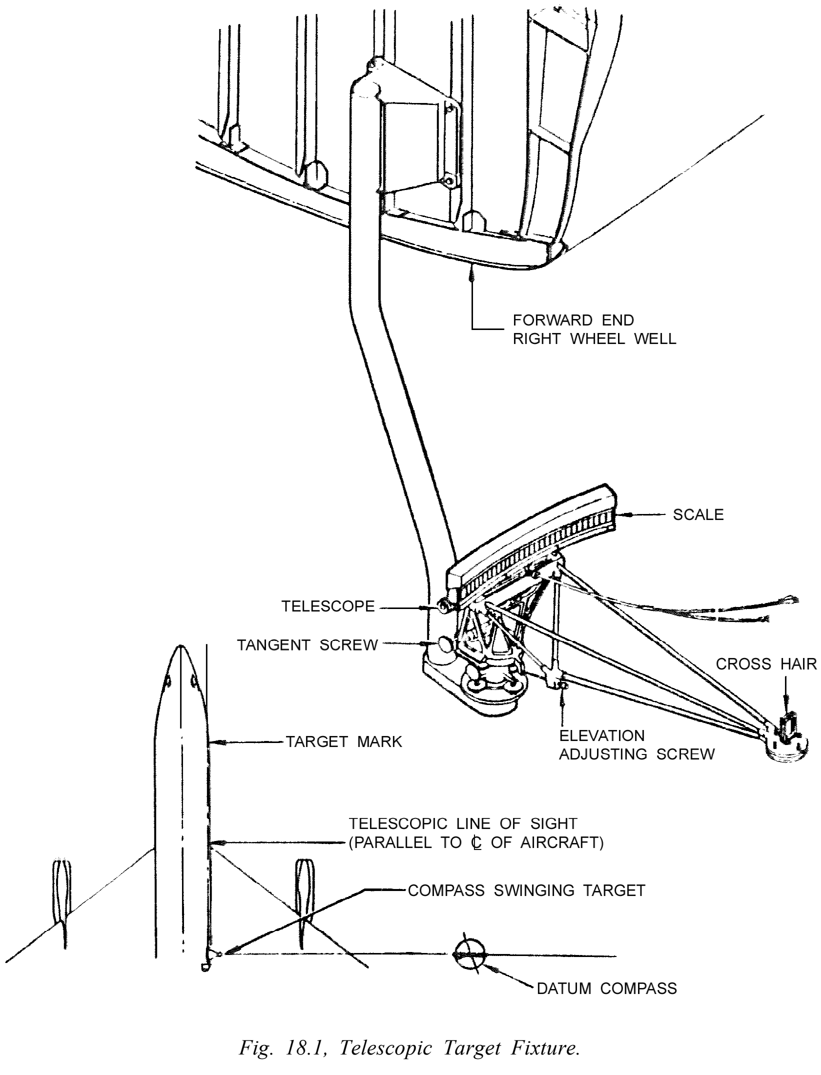



*Fig. 18.1, Telescopic Target Fixture.*

b) Another example of a target fixture is shown in Fig. 18.2. In this case the fixture is attached to the underside of the front fuselage, and its telescope is sighted on a target ball mounted on a cable, the ends of which are secured to the up-lock rollers of each main landing gear strut. When the cable is in position, the ball is situated to the left of the aircraft centre line. Magnetic heading of the aircraft is determined in a similar manner to that described in (a).

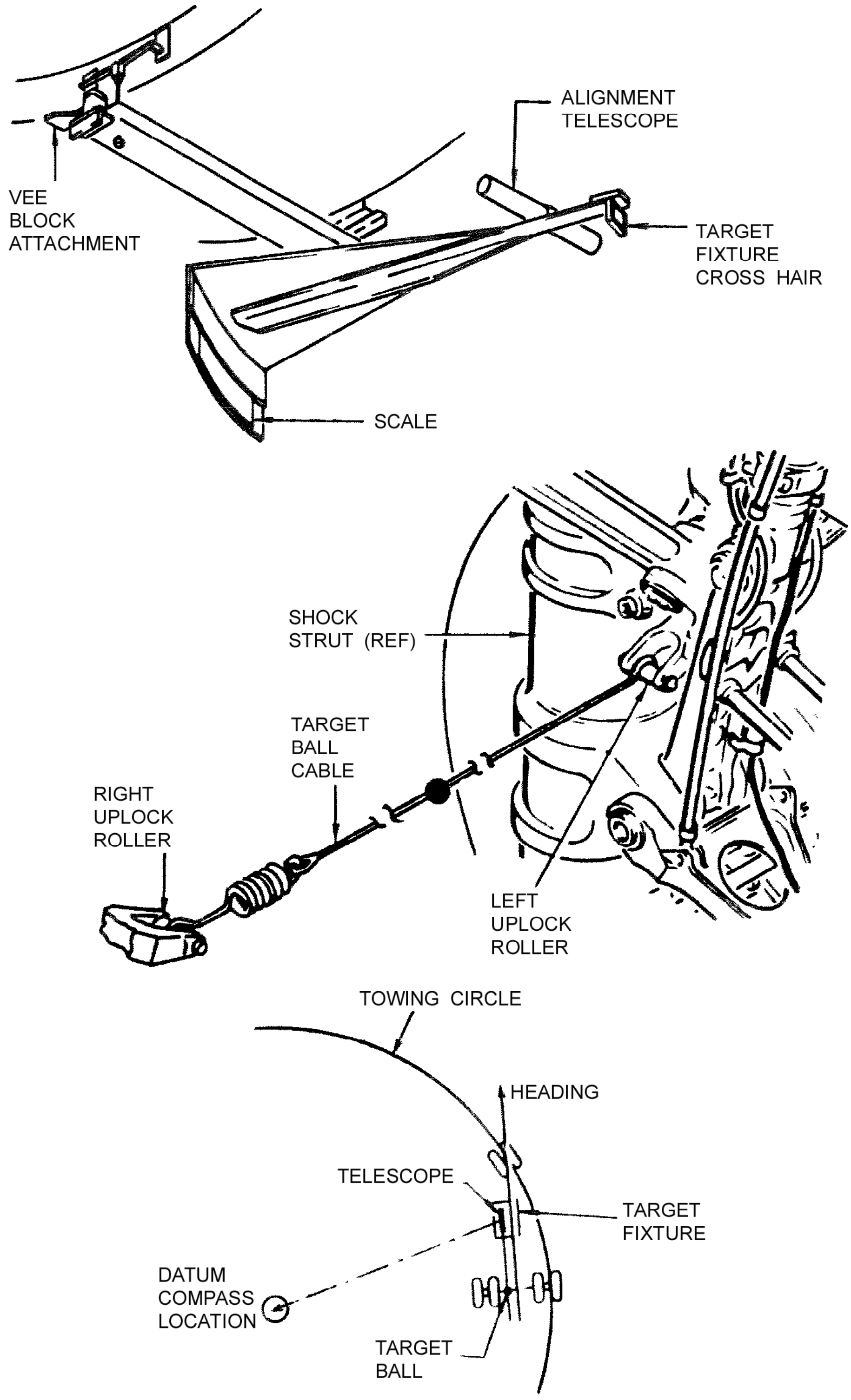



*Fig. 18.2, Target fixture and ball.*

**Plumb Line Sighting**

In this method, the longitudinal axis of an aircraft is indicated by the alignment of two plumb bobs suspend individually from fore and aft points at the underside of the fuselage. The use of the plumb bobs as a means of sighting depends on the compass swinging procedure adopted. If a swing carried out on a base on which cardinal and quadrantal heading lines are marked out, the aircraft heading is determined by the position of the plumb bobs with respect to the marked lines.

In the case of an "electrical" swinging procedure, the plumb bobs are used only for ensuring that aircraft is aligned with the North-South line marked out on base. Malalignment should not exceed a specified amount, usually 1 degree, and this is calculated from the nomograph comprising three scales corresponding to the separation between plumb bobs, to plumb bob displacement from the North-South line, and to aircraft malalignment. An example based on an aircraft in which the plumb bob suspension points are 76 ft 2 in. apart is shown in Fig. 18.3

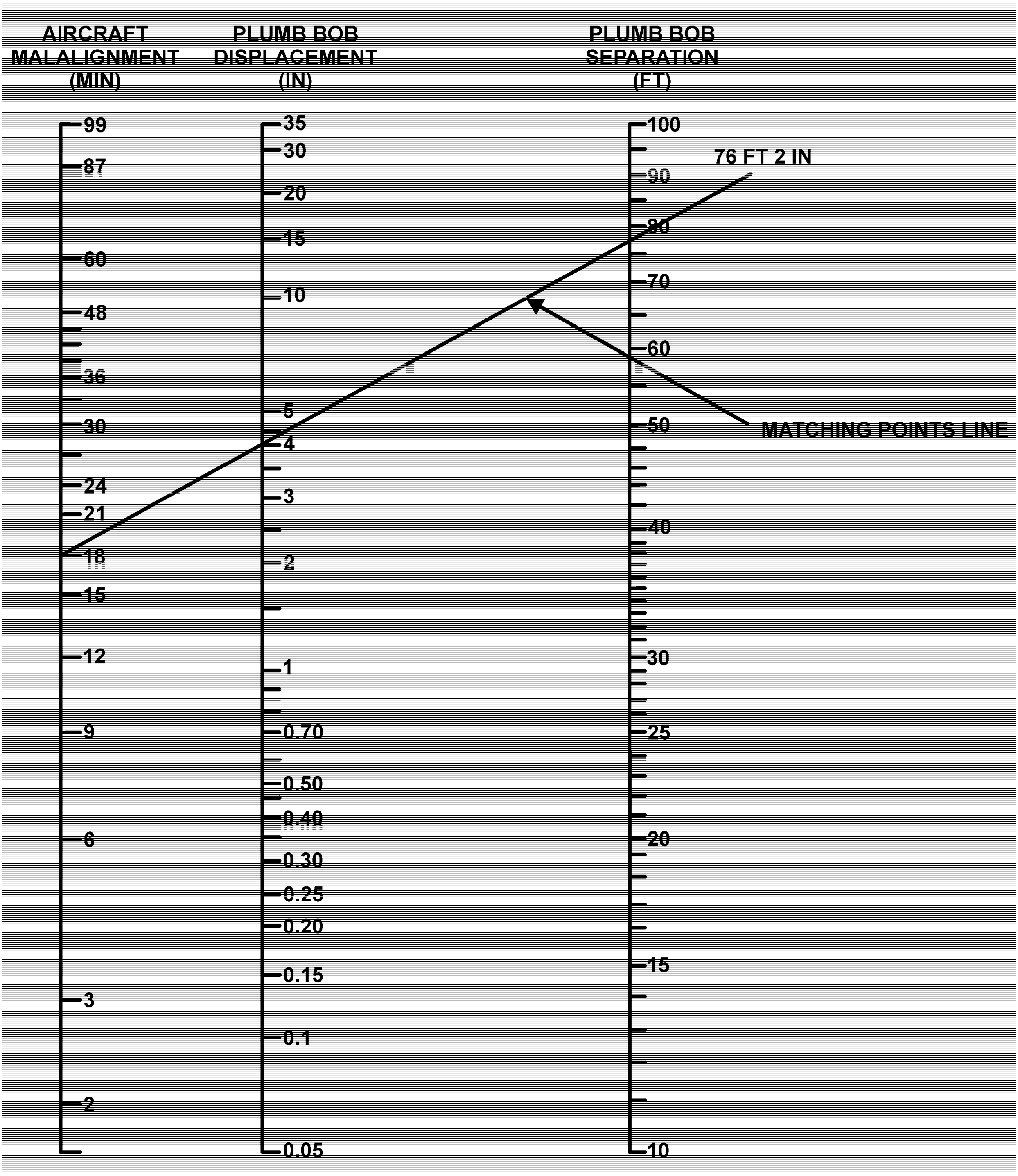



*Fig. 18.3, Aircraft Malalignment Nomograph.*

When the aircraft is positioned over the North-South line, with its nose pointing North, the points where the plumb bobs come to rest are marked on the ground and the lateral displacement of each from the edge of the line is measured. If the displacement is West of the line it is negative, and if it is East it is positive. The algebraic difference between the two values is then obtained as a reference point on the plumb bob displacement scale, and aircraft malalignment is read off the appropriate scale at the point intersected by a line projected through the known points on the plumb bob separation and projection scales. Thus, in the example shown, the known pints are 76 ft.2 in and 4 in., and a projected line intersects the malalignment scale at a value of 18 minutes. In the event that malalignment is greater than the maximum specified for the aircraft, the aircraft should be repositioned.

**Conventional Swinging Procedure**

The term "conventional" is used here to signify a procedure in which the magnetic heading reference datum is obtained, either from, a compass base with heading alignment marks painted on it, or from a datum compass. In the former case, and with the aid of plumb line sighting, an aircraft can be aligned precisely on the cardinal and quadrantal headings. When using a datum compass, such precise alignment is not essential since the compass is always positioned to sight on the aircraft. It is usual, therefore, for positioning of both aircraft and datum compass to be within certain permissible limits. Limits commonly specified are within the range ± 3 degrees to ± 5 degrees. A typical swinging sequence is outlined in the following paragraphs.

When the aircraft has been towed onto the base, the appropriate sighting equipment should be fitted and the aircraft positioned so that it is heading North.

External power supply should be connected to the aircraft, and after the compass systems have been energized and their gyros allowed to run up to normal operating speed, carry out the preliminary checks specified in the appropriate Maintenance Manuals. The following checks are typical of those normally required for systems in current use:

a) Synchronizing of heading indicators against annunciator devices to ensure that magnetic monitoring by flux detector units, has ceased before taking readings. In certain types of compass system this check is effected by plugging a centre-zero milliammeter into "monitoring current" socket provided in a unit (e.g. and amplifier) of the system. If the gyro has no drift, monitoring has ceased with the meter oscillations are evenly balanced about the indicator "null" position within the tolerances specified for the particular system. In some cases, a monitoring meter forms part of a compass indicator, thereby obviating the need for a centre zero milliammeter.
b) Slaving of compass system indicators and associated system indicators, e.g. radio magnetic indicators.
c) Check heading signals to auto-pilot and other associated navigational systems after selection of appropriate system switch positions.
d) Operational check of power failure warning and other indicating flags on all heading indicators.
e) Drift rate check of gyros.
f) Setting of deviation compensators to their neutral positions. This is normally only done during an initial swing procedure, and whenever a new flux detector unit or a deviation compensator is installed. If a compensator is of the permanent magnet type, the slots of the adjustment screws should be aligned with datum marks on the compensator body. In the case of a potentiometric type compensator, the potentiometers should be adjusted until a "null" position, as indicated by a test meter plugged into the compensator is obtained. If a new flux detector unit has been installed its index error scale should also be aligned at the zero datum.

With the aircraft still heading North and compass indicators synchronized, the deviation, i.e. the difference between the indicated reading and the magnetic datum, should be calculated. The sign is plus or minus according to whether it is necessary to add deviations to, or subtract them from, the compass readings in order to obtain the magnetic heading.

Position the aircraft so that it is heading East and after allowing the compass indicators to synchronize, the deviation should be calculated and its sign determined.

Position the aircraft so that it is heading South and after calculating the deviation, determine the Coefficient "C" error by calculating the average of the algebraic difference between the deviations on the North and South headings. The sign of the coefficient should then be changed and added algebraically to the indicated readings to establish the corrected heading indicators. The Coefficient "C" section of the appropriate deviation compensation devices should then be adjusted until, with the indicators synchronized, the corrected heading is indicated.

Position the aircraft so that it is heading West and, after calculating the deviation, determine the Coefficient "B" error by calculating the average of the algebraic difference between the deviations on the East and West headings. The corrected headings indications are established and applied in a similar manner to that described above except that adjustments are made at the Coefficient "B" section of the appropriate compensation devices.

On completion of the Coefficient "B" correction the aircraft should be positioned successively on the cardinal and quadrantal headings (starting at West) in order to calculate the Coefficient "A" error. The error is calculated by taking the average of the algebraic sum of the deviations on eight headings, and is corrected in the manner appropriate to the compass system installed, After applying the Coefficient "A" correction, the compass indicator readings on the cardinal and quadrantal headings should again be noted and the residual deviations should be recorded.

All observed readings and associated deviation calculations obtained during swinging should be recorded on properly prepared record forms, and a record of the swing should be entered and certified in the aircraft log book. A record should also be made of compensator settings. A deviation or "steer by" card should also be complied from recorded residual deviations, to show the readings which a compass must indicate in order that the aircraft may be flown on correct magnetic headings. The deviations should be related to standard headings at intervals of 45 degrees; in some cases an interval if 30 degrees may be specified. Details should be given on the back of the card to indicate aircraft type and registration, compass type, place and date of swing, signature and authority of the complier. On completion, the card should be displayed in the appropriate holder in the aircraft cockpit.

**"ELECTRICAL" SWINGING PROCEDURE**

An "electrical" swinging procedure is one in which the earth's magnetic field is simulated by electrical signals in such a ways that it unnecessary to rotate the aircraft onto the various headings as in the conventional forms of swinging. The aircraft is positioned heading North with its fore-and aft axis coincident with a North-South line marked on the selected compass base, and with its flux detector units positioned over marked location datum points. The electrical signals are in the form of varying d.c. voltage, the values of which are determined during the appropriate compass base survey procedure and also during the swinging procedure. The signals are designated as E1 and E2 voltages and are applied respectively to the A-leg pick-off coil, and the B leg and C-leg pick off coils of the flux detector sensing elements, by adjusting the controls of a console unit forming part of a special calibrator set which can be connected into the flux detector circuit. The electromagnetic fields produced by the E1 and E2 voltages, alter the effective magnitude and direction of the earth's field passing through the legs of the detector sensing element, resulting in a magnetic vector which rotates to various headings, thereby simulating rotation of the aircraft and detector unit relative to the earth's field. These simulated headings are compared with the actual headings indicated by the aircraft compass system to determine the deviation errors to be compensated.

**COMPASS CALIBRATOR SET**

It is beyond the scope here to go into the complete details of the compass calibrator set and its use in the "electrical" swinging procedure, reference should, therefore, always be made to the relevant manufacturer's operation manual. The information given in the following paragraphs serves only as general outline.

a) A compass calibrator set consists of the major components shown in Figure 18.4 and 18.5. The magnetic field monitor is a theodolite with a 22x telescope and a magnetic sensing element, and in conjunction with the console unit. It measures the strength, and determines the direction, of the horizontal component of the earth's field it is used for the magnetic survey of areas required for swinging and also to monitor changes in magnetic conditions during a compass swing. The magnetic sensing element is similar to that employed in remote indicating compass flux detector unit, except that it is non-pendulous.

b) The turntable is also a theodolite but without a telescope and magnetic sensing element. It is used for calibrating and aligning certain types of flux detector unit with magnetic North, and for determining the index or Coefficient "A" error before installing a unit in the aircraft.

c) The console unit is central control unit of the calibrator and contains all the switches, controls and indicators for programming the E1 and E2 voltage signals and for determining the errors in aircraft compass system indications to be compensated. The console also provides interconnection of the magnetic field monitor, turntable and power supply unit. The power supply unit is a solid-state inverter which converts 28V d.c. input into a 115V 400 Hz a.c. output required to operate the calibrator.

d) The optical alignment unit fig.18.5., consists of a fixed-focus 8x telescope and appropriate adjustment devices, and is used for aligning certain types of flux detector unit in an aircraft during its transfer from the turntable, thereby ensuring that the index error compensation is maintained.

**SWINGING PROCEDURE**

The procedure for carrying out an "electrical" swing depends primarily on whether a flux detector unit is of the pre-indexed type, or of the master type (precalibrated and pre-indexed). Basically, however, the procedure consists of the following sequence of operations:-

**a) Check On The Direction Of Magnetic North**

This is done to determine whether there has been any shift from that obtained when the base survey carried out. The check is carried out by sighting the calibrator monitor, from its location point on the base, on the predetermined reference target, and then determining errors on the cardinal headings and thus obtain an area compensation value for setting on the console control unit.

**b) Magnetic Alignment of Detector Unit**

The purpose of this operation is to check the Index or Coefficient "A" error, and thereby the amount by which the detector unit is to be offset in its mounting with respect to magnetic North. If the detector units are of the pre-indexed type, the check is done with the unit mounted on the calibrator turntable at its location on the base, and before the aircraft is towed onto the compass base. In the case of master type detector units, the aircraft can be towed onto the base at the outset of the swing procedure, since the units, being pre-calibrated on the four cardinal headings for a specific compass system and aircraft installation, are already installed and the use of the turntable is thereby eliminated.

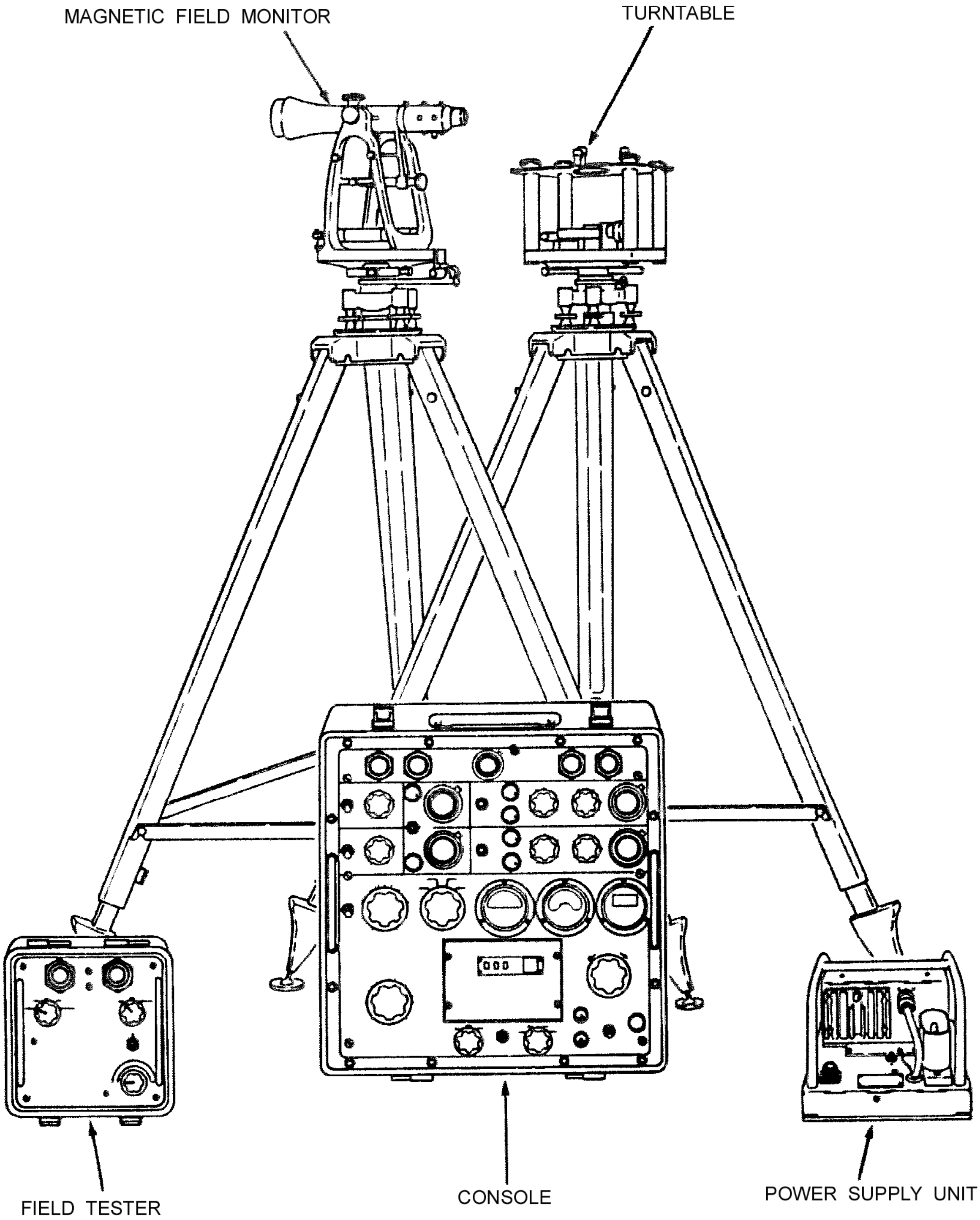



*Fig. 18.4, Compass Calibrator Set.*

**c) E1 and E2 Voltages Check**

The purpose of this check is to determine the voltages required to simulate the earth's field effects which would be obtained if the aircraft and its detector units were rotated onto various headings. The check also determines the adjustments which are necessary to compensate for one-cycle errors, i.e. Coefficients "B" and "C", during the compass swing operation. The check applied to both pre-indexed and master type flux detector units, except that in the former case, voltage values are obtained by selecting headings on the calibrator turntable, while for master units selections are made on the calibrator monitor.

**d) Determination of Crosstalk Errors**

These errors are measured at headings of 90, 180 and 270 degrees. Crosstalk error does not occur at 0 degrees since no voltages are applied to the flux detector unit to simulate this heading.

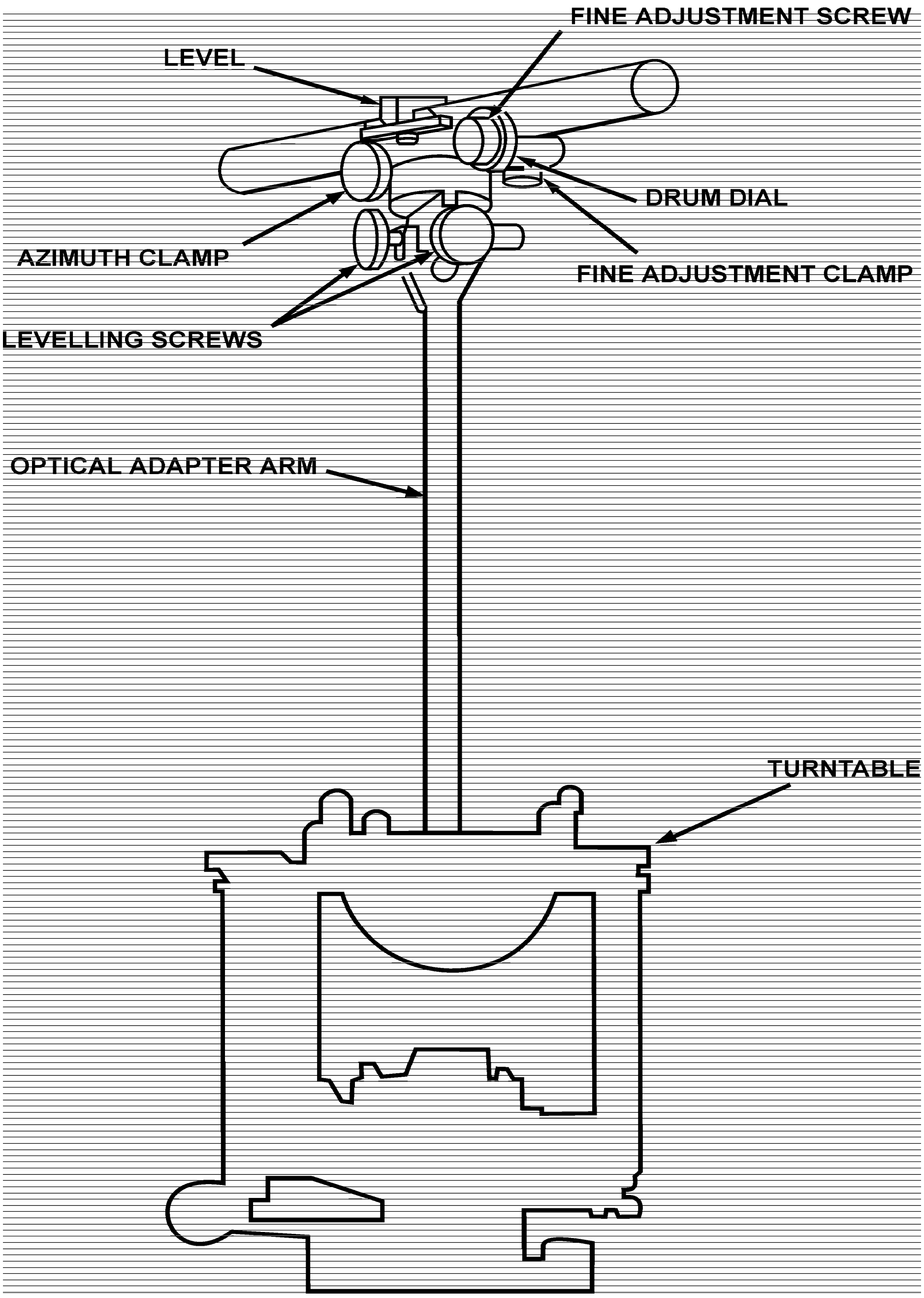



*Fig. 18.5, Optical Alignment Unit.*

**e) Optical Transfer Of Flux Detector Unit**

The operation is applicable only to flux detector units which have not been pre-indexed or pre-calibrated. It is carried out by means of the calibrator optical alignment unit, and is necessary in order to ensure that a detector unit will be installed in the same position with respect to magnetic North, as that determined by the calibrator turntable magnetic alignment check. The aircraft is towed onto the base and positioned so that not only is the longitudinal axis coincident with the North-South line with the nose pointing North, but also the flux detector unit access location is directly over that of the calibrator turnable. The optical alignment unit is then attached to the detector unit and a reference bearing is obtained by sighting the telescope on a target which is at least half-mile distant from the base. The values of aircraft malalignment, and flux detector unit index error, are then set into the optical alignment unit, and the detector unit, with equipment attached, is transferred to its mounting bracket in the aircraft and is rotated until the telescope is again aligned with the distant reference target. At this setting the detector unit is secured to its mounting bracket and electrically connected to its compass systems. The optical alignment unit is then removed.

**f) Compass Swinging**

This operation, which applies to all compass systems irrespective of the type of flux detector unit employed, is the one in which compensations are made for one cycle (Coefficients "B" and "C") and two cycle (transmission) errors. Before compensating, a final check on E1 and E2 voltage values should be made and adjustments should, where necessary, be carried out to allow for any possible changes in the earth's field strength. On completion of all adjustments, a final swing through increments of 15 degrees should be carried out.

**SWINGING BY INERTIAL NAVIGATION SYSTEMS**

With the introduction of Inertial Navigation System (INS) into certain types of public transport aircraft, the use of system display unit heading information, as the datum for compass swinging, became a possibility. The swinging procedure, although basically similar to that employing an external magnetic datum, does however, require that the effects of diurnal changes in local magnetic variation be taken into account in order to minimize compass deviation errors.

Variation is the horizontal angle between the true an magnetic meridians, and in the United Kingdom it normally increases Westerly during the morning and decreases during the late afternoon. These diurnal changes, as they are called, are in the order of 0.1 degrees in the winter, and 0.3 degrees in the summer. During periods of high sunspot activity however, the changes become random and may increase in amplitude to 0.5 degrees, and have been known to be as large as one degree. In the case of a compass swing using an external magnetic reference datum, the diurnal changes do not affect the swing since both the datum compass and the aircraft compass are affected by the same changes and the correct deviations are calculated. Furthermore, as long as the compass base has a constant value of variation over it, an accurate compass swing can be carried out irrespective of the value of local variation. In using INS heading information as a datum, diurnal changes will, however, affect only the aircraft compass, thereby giving rise to false deviation errors which require the application of variation corrections. An outline of a method of applying the corrections, based on that prescribed for a particular type of aircraft using three inertial navigation systems, is given in the following paragraphs:

a) The aircraft is positioned within two degrees of North using the readings of the INS display units, and the average of the three readings is noted. The aircraft's magnetic heading is then determined by applying the value for the local magnetic variation to the average of the display unit readings. For Westerly variation the value is added and for Easterly variation it is subtracted.
b) The heading indications of the compass systems are then noted, after allowing each system to synchronize and the deviations of each system on the North heading are calculated.
c) The foregoing procedure is repeated on the other three cardinal headings, and the deviation coefficients are calculated and compensated in the manner prescribed for the particular compass system.

Although the method takes into account local variation on all cardinal headings, it should be noted that it is still possible for false deviations to occur as a result of diurnal changes taking place, for example, during the time required for compass systems to synchronize. Unless continuous calculations are made, or the aircraft headings are checked with the aid of an external datum compass, it is unlikely for local variation to be known accurately for the duration of swing. Serious consideration should, therefore be given to the maximum compass error which might occur during an INS datum swing compared with one using an external magnetic datum, and whether such error can be accepted for the aircraft compass systems concerned.

**OCCASIONS FOR COMPASS SWINGING**

The swinging procedures described in normally be carried out after installation of a complete compass system, and whenever standard type, or pre-indexed type, flux detector units are changed. Changing of a master type detector unit does not usually downgrade the performance of its associated system unless alignment of the unit in its mounting bracket or alignment of the bracket itself, has been altered. Normally, a complete swinging procedure should also be carried out after a deviation compensator device has been changed although in some systems this may not be necessary provided that compensating voltage settings are properly transferred to the replacement unit. On all other occasions it is sufficient only to carry out a check swing by placing the aircraft on four headings 90 degrees apart and comparing any deviations with those recorded on the previous calibration swing. If there is any difference between these deviations a complete swinging procedure should be carried out. Occasions for a check swing are as follows :

a. After a check inspection if required by the approved Maintenance Schedule or at any time that the accuracy of a system is in doubt.
b. After any modification, repair or major replacement involving magnetic material.
c. After any modification or repair to wing tips or vertical stabilizers in the vicinity of flux detector units.

d. Whenever a compass has been subjected to shock, e.g. after a heavy landing.
e. After the aircraft has passed through a severe electrical storm or has been struck by lightning.
f. Whenever the aircraft has been subjected to a magnetic crack detection examination.
g. Whenever the sphere of operation of the aircraft is changed to one of different magnetic latitude.
h. Whenever a significant change is made to the electrical or radio installation, particularly to circuits in the vicinity of flux detector units.
j. Whenever a freight load is likely to cause magnetic influence and thereby affect compass system readings.
k. After the aircraft has been in long term storage.

■■■

# CHAPTER : 19
# ENGINE INSTRUMENTS

**GENERAL**
Many of the instruments utilise electrical elements based on the principle of measuring the required variables at source and transmitting the data through the medium of a synchronous signal link system. Various types of synchronous data transmission system may be used and as some are applied to certain of the instruments to be described, an outline of their general construction and operation is given at this stage to avoid repetition.

**SYNCHRONOUS DATA TRANSMISSION SYSTEMS**
All the systems consist of two principal components : a transmitting element and a receiving element, electrically interconnected through signal lines and supplied with power from aircraft electrical system. Depending on the synchronous link adopted the power supply required may either be direct current or alternating current. The most common form of d.c synchronous system is the "Desynn" of which there are three circuit variations; the application of each circuit being governed by the variables to be measured.

**Basic Desynn Circuit**
This circuit is applied to engine systems in which the position of a mechanical component is required to be known, e.g. a fuel trim actuator or a float mechanism of a fuel quantity indicator.

The transmitter unit consists of a circular wire-wound resistance or toroidal tapped at three points $120^0$ apart, and a brush assembly made up two diametrically opposed contact arms suitably insulated from each other. The brush assembly rotates in contact with the toroid and serves to lead the d.c supply to the system. Rotation of the brush assembly is effected by a slotted arm which engauges with a crank pin connected through a mechanical linkage to the appropriate mechanical component. The type of linkage varies, but in a typical application to component position indication (e.g. fuel trim actuator position) it consists of a lever arm and spring-controlled gearing. The lever arm normally operates through $60^o$ against a control spring, and the gear ratio between the arm and the crank pin may either be 3:1 or 6:1 so that the brush assembly can move through $180^o$ or $360^0$. "Limit Stops" lever arm movement to $70^o$.

The receiving or indicating part of instrument is made up of a two-pole permanent magnet rotor pivoted inside a soft-iron stator carrying a three-phase star connected winding supplied from the three tappings of the transmitter toroid. A pointer is attached directly to the rotor spindle and rotates over a scale calibrated in the units appropriate to the required measurements.

When the brush assembly rotates over the toroid, the current flow to the indicator stator winding varies in magnitude and direction. The current is distributed through the windings and produces a magnetic field which rotates in synchronism with the brush assembly. The magnet aligns itself with this field, thus, carrying the pointer to positions indicating the amount of brush movement. When the power supply to the system is interrupted the pointer is returned to an off-scale position by a weak pull-off magnet which attracts the main magnet rotor.

**Micro-Desynn Circuit**
The Micro-Desynn circuit is applied to systems in which less power and very small displacements of a primary actuating system is available, e.g. in the measurement of fuel or oil pressure. The circuit is a development of the basic one and the principle remains the same. In place of the toroidal resistance, however, two resistance coils are used and the brushes are arranged to move together over the whole length of their respective coils and through an angle of about $60^o$. This movement, combined with the method of electrically tapping the coils, corresponds to one revolution of the contacts of a toroidal transmitter and results in $330^o$ movement of the indicator pointer. The brushes are rotated by means of a push-rod and rocking shaft assembly coupled to a bellows type of pressure-sensing element. In applications of the circuit to liquid quantity measurement, the brushes are rotated by a mechanically coupled float mechanism. With the exception of the dial calibration markings, the receiving or indicating element is similar to that employed in the basic Desynn circuit.

**Slab-Desynn circuit**
In both the basic and micro circuits, the voltage at each of the transmitter tappings, although varying proportionally with the brush movements, produces a magnetic field in the indicator stator which causes a light cyclic error. This deviations, plus frictional losses and contact wear, are characteristics of both circuits which can be taken into account during initial calibration. For some applications it is necessary for the effects of these characteristics to be minimised within the circuit itself and for this reason the Slab-Desynn circuit was developed.

The resistance of the transmitter element is wound on a flat former, hence the term 'slab', and the power supply is connected at either end. This arrangement produces a uniform potential gradient. The pick-off assembly consists of three brushes spaced at 120º and is equivalent to the three tapping on the other types. As the assembly rotates, voltage are distributed to the indicator stator which vary according to a sine law. Another feature of this arrangement is that as each contact carries less current wear is reduced.

**A.C. Synchronous Systems**

These systems which are generally known as 'synchros' are manufactured under a variety of trade names such as "Selsyn", "Autosyn" and "Asynn" the names being contractions of the functional terms adopted, e.g. "Selsyn" is a contraction of 'self synchronous'. The operating principle of these systems is basically the same, each consisting of electrical transmitter and receiving elements. However, unlike d.c. systems, both elements are similar in construction and employ a two-pole single phase rotor free to rotate inside a three-phase wound stator. The stator windings of each element are interconnected and in most instrument applications the rotors are supplied with alternating current (26-volts or 115-volts) via slip rings.

When alternating current is supplied to the rotors a definite combination of voltages is induced in both stator windings. If both rotors are in the same position in relation to their respective stators both sets of stator voltages will be equally opposed and no current will flow in the coils. Thus, there is no magnetic field torque and the rotors remain in alignment. When the transmitter rotor is moved the induced voltages are unbalanced and currents flow in the windings of both stators producing magnetic fields which turn the receiver synchro rotor to the same position as that of the transmitter, restoring a balanced condition. There is a tendency for the transmitter synchro rotor to be turned within its stator but as it is mechanically coupled to the appropriate measuring element, it is prevented from doing so and the receiver synchro is made to follow the transmitter.

**FUEL QUANTITY GAUGES**

The measurement of the quantity of fuel in the tanks of an aircraft fuel system is an essential requirement, and in conjunction with measurements of the rate at which the fuel flows to the engine or engines, permits an aircraft to be flown at maximum efficiency compatible with its specified operating conditions. Furthermore, both measurements enable a pilot or engineer to quickly assess the remaining flight time and also to make comparisons between present engine performance and past or calculated performance.

Fuel-quantity indicating systems vary in operating principle and construction, the application of any one method being governed by the type of aircraft and its fuel system. Two principal methods currently applied utilize the principle of electrical signal transmission from units located inside the fuel tanks. In one method, mainly employed in the fuel systems of small and light aircraft, the tank unit consist of a mechanical float assembly which controls an electrical resistance unit and varies the current flow to the indicating element. The second method , employed in high-performance aircraft fuel systems, measures fuel quantity in terms of electrical capacitance and provides a more accurate system of fuel gauging.

Fuel-flow measuring systems also vary in operating principal and construction by principally they consist of two units: a transmitter or meter, and an indicator. The transmitter is connected at the delivery side of the fuel system, and is an electromechanical device which produces an electrical output signal proportional to the flow rate which is indicated in either volumetric or mass units. In some systems an intermediate amplifier/computer is included to calculate a fuel-flow/ time ratio and also to transmit signals to an indicator which presents integrated flow rate and fuel consumed information.

**Direct Reading Fuel Quantity Gauge**

Consists of a float made of cork which raises or lowers with the level of fuel inside the tank. Float is attached to float arm which drives a gear or pinion and pinion is connected to a pointer which moves over a calibrated dial and indicates the quantity of fuel in the tank in gallons.

**Float-Type Fuel-Quantity Indicating Systems (Electrical Type)**

The components of a float-type system are shown in schematically in Fig. 19.1, together with the methods of transmitting electrical signals.

The float may be of cork specially treated to prevent fuel absorption, or it may be in the form of a lightweight metal cylinder suitably sealed. The float is attached to an arm pivoted to permit angular movement which is transmitted to an electrical element consisting of either a wiper arm and potentiometer, or a Desynn type of transmitter. As changes in fuel level take place the float arm moves through certain angles and positions the wiper arm or brushes to vary the resistance and flow of direct current to the indicator. As result of the variations in current flow a moving coil or rotor within the indicator is deflected to position a pointer over the scale calibrated in gallons.

**Capacitance-Type Fuel-Gauge System**

In its basic form, a capacitance-type fuel-gauge system consists of a variable capacitor located in the fuel tank, an amplifier and an indicator. The complete circuit forms an electrical bridge which is continuously being rebalanced as a result of differences between the capacitances of the tank capacitor and a reference capacitor. The signal produced is amplified to operate a motor, which positions a pointer to indicate the capacitance change of the tank capacitor and thus the change in fuel quantity.

Before going into the operating details of such a system, however, it is first necessary to discuss some of the fundamental principles of capacitance and its effects in electrical circuits.

**Electrical Capacitance**

Whenever a potential difference is applied across two conducting surfaces separated by non-conducting medium, called a dielectric, they have the property of storing an electric charge; this property is known as capacitance.

The flow of a momentary current into a capacitor establishes a potential difference across its plates. Since the dielectric contains no free electrons the current cannot flow though it, but the potential difference sets up a state of stress in the atoms comprising it . For example, in the circuit shown in Fig. 19.2, when the switch is placed in position 1 rush of electrons, known as the charging current, takes place from plate A through the battery to plate B and ceases when the potential difference between the plates is equal to that of the battery.

When the switch is opened, the plates remain positively and negatively charged since the atoms of plate A have lost electrons While those at plate B have a surplus. Thus, electrical energy is stored in the electric field between the plates of the capacitor.

Placing the switch in position 2 causes the plates to be short-circuited and the surplus electrons at plate B rush back to plate A until the atoms of both plates are electrically neutral and no potential difference exists between them. This discharging current is in the reverse direction to the charging current, as shown in Fig. 19.2.

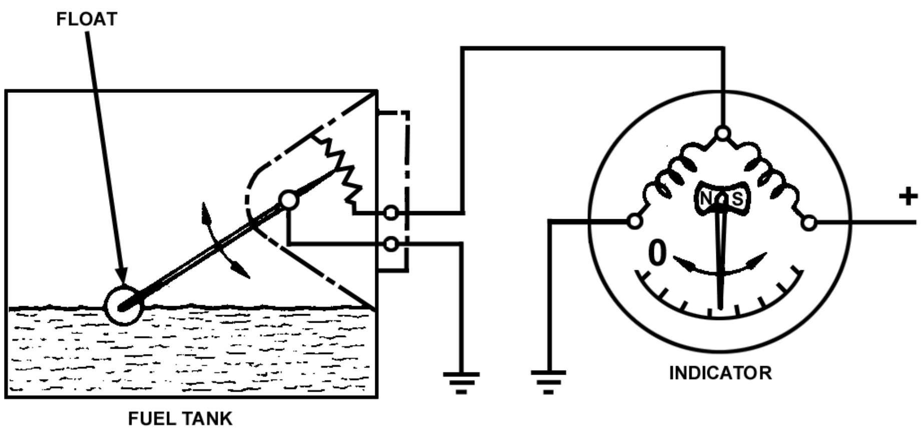



*Fig. 19.1, Simple float type of fuel quantity indicator.*

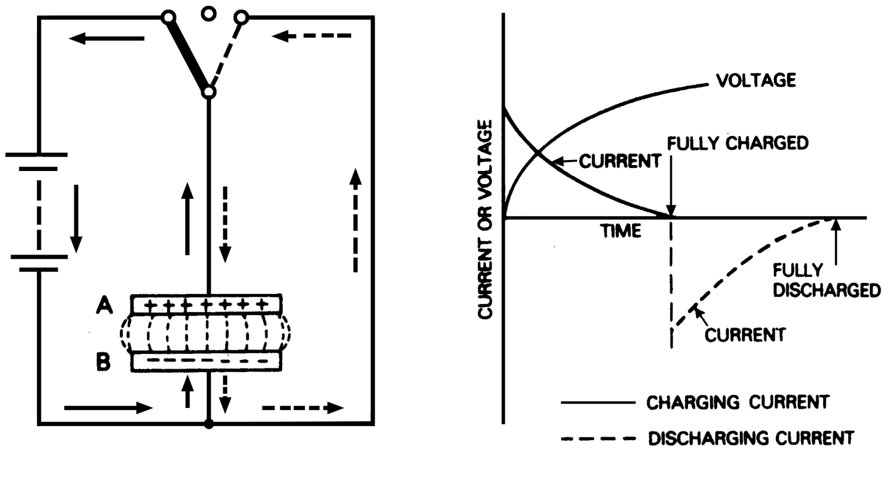



*Fig. 19.2, Charging and discharging of a capacitor.*

**Units of Capacitance**

The capacitance or "electron-holding ability" of a capacitor is the ratio between the charge and the potential difference between the plates and is expressed in farads, one farad representing the ability of a capacitor to hold a charge of one coulomb ($6.24 \times 10^{18}$ electrons) which raises the potential difference between its plates by one volt.

Since the farad is generally too large for practical work, a sub-multiple of it is normally used called the microfarad

($1 \mu F = 10^{-6} F$). In the application of the capacitor principle to fuel gauge systems, an even smaller unit, the picofarad ($1\ pF = 10^{-12}F$) is the standard unit of measurement.

**Factors on Which Capacitance Depends**

The capacitance of a parallel-plate capacitor depends on the area, a, of the plates, the distance, d, between the plates, and the capacitance , $\varepsilon_{\alpha}$ of a unit cube of the dielectric material between the plates :

$$C = \varepsilon_a (a \backslash d) \text{ farads}$$

The units of $\varepsilon_a$ is the farad per meter, so that, a must be expressed in square meters, and d in meters; $\varepsilon_a$ is called the absolute permittivity of the dielectric.

It is usual to quote permittivities relative to that of a vacuum, whose permittivity, $\varepsilon_o$, is $1/(4 \pi \times 9 \times 10^9$ F)/m. Relative permittivity, ε, is also called dielectric constant and is often denoted by K. In terms of relative permittivity,

$$C = [\{k / (4 \pi \times 9 \times 10^9\} (a/d)]$$

The relative permittivity of air at standard temperature and pressure is 1.00059, which for practical purpose & may be taken as 1.0. Then, for example,

$$C_{water} \backslash C_{air} = K_{water} \backslash K_{air} = K_{water} \text{ (very closely)}$$

i.e. , K is the ratio of the capacitance of a capacitor with a given dielectric to its capacitance with air between its plates.

The relative permittivity of some pertinent substances are as follows:

| | |
|---|---|
| Air | 1.00059 |
| Water | 81.07 |
| Water vapour | 1.007 |
| Aviation gasolene | 1.95 |
| Aviation Kerosene | 2.10 |

**Capacitors in Series And Parallel**

The total capacitance of capacitors connected in series or parallel is obtained from formulae similar to those for calculating total resistance but which are applied in the opposite manner. Thus, for capacitors connected in series, the total capacitance is given by

$$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots..$$

and for capacitors connected in parallel,

$$C_T = C_1 + C_2 + C_3 + \ldots..$$

This is because the addition of capacitors in a circuit increases the plate area which, as already stated, is one of the factors on which capacitance depends.

**Capacitors In Alternating-Current Circuits**

When direct current is applied to a capacitor there is, apart from the initial charging current, no current flow through the capacitor. In applying the capacitance principle to fuel-gauge systems, however, a flow of current is necessary to make the indicator respond to the changes in capacitance arising from changes in fuel quantity. This is accomplished by supplying the capacitance-type tank units with an alternating voltage, because whenever such voltage across a capacitor changes electrons flow toward and away from it without crossing the plates and a resultant current flows which, at any instant, depends on the rate of change of voltage.

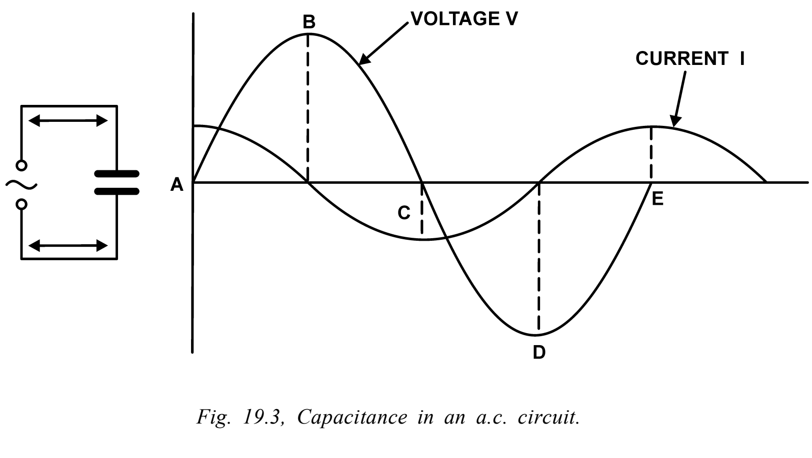



*Fig. 19.3, Capacitance in an a.c. circuit.*

Fig 19.3 shows a capacitor connected to an alternating-current source. It will be observed from the graph that, as the voltage (V) rises rapidly at A a large current (I) will flow into the capacitor to charge it up. As the voltage increases towards B, however, the current decreases until at B, when the voltage is steady at some maximum value for a brief instant, the current has decreased to zero. From B to C the voltage decreases, the capacitor discharges and the current flows in the opposite direction, being a maximum at C, where the voltage is zero. From C to D the capacitor is charged in the opposite direction and the current flows in the same direction as the voltage but reaches zero at D, where the voltage is at some maximum value in the opposite direction. From D to E the capacitor again discharges. Thus, a charge and current flow in and out of the capacitor occurs every half-cycle, the current leading the voltage by 90°.

The ratio between the voltage V and the current I is termed the capacitive reactance, meaning the opposition or resistance a capacitor offers to the flow of alternating current.

**Basic Gauge System**

For fuel quantity measurement, the capacitors to be installed in the tanks must differ in construction from those normally employed in electrical equipment. The plates therefore take the form of two tubes mounted concentrically with a narrow air space between them, and extending the full depth of a fuel tank. Constructed in this manner, two of the factors on which capacitance depends are fixed, while the third factor, dielectric constant, is variable since the medium between tubes is made up fuel and air. The manner in which changes in capacitance due to fuel and air take place is illustrated in the Fig. 19.4. and is described in the following paragraphs.

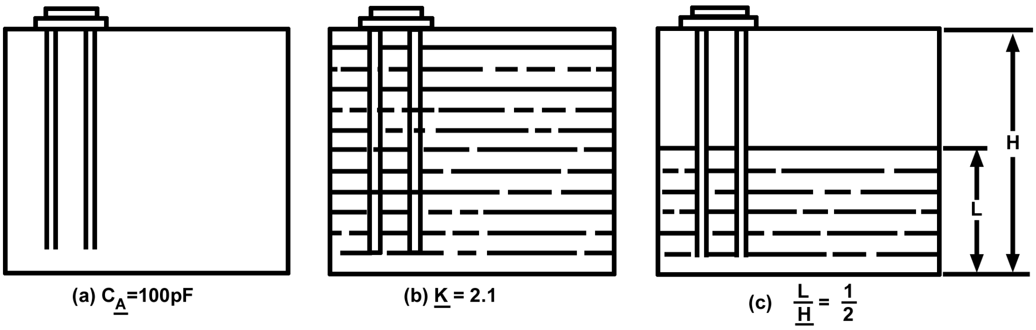



*Fig. 19.4, Changes in capacitance due to fuel and air.*

At (a) a tank capacitor is fitted in an empty fuel tank and its capacitance in air is 100 pF, represented by $C_A$.

At (b) the tank is filled with a fuel having a K value of 2.1 so that the capacitor is completely immersed. As stated K is equal to the ratio of capacitance using a given dielectric (in this case $C_T$) to that using air; therefore.

$$K = C_T \backslash C_A \quad (1)$$

From eqn. (1) $C_T = C_A K$, and it is thus clear that the capacitance of the tank unit at (b) is equal to 100 x 2.1 i.e. 210 pF. The increase of 110 pF is added capacitance due to the fuel and may be represented by $C_F$. The tank unit may therefore be represented electrically by two capacitors in parallel and of a total capacitance

$$C_T = C_A + C_F \quad (2)$$

In Fig.19.4 (c), the tank is only half full and so the total capacitance is 100 + 55, or 155 pF. The added capacitance due to fuel is determined as follows. By transposing eqn (2), $C_F = C_T - C_A$ and by substituting $C_A K$ for $C_T$ we obtain $C_F = C_A K - C_A$ which may be simplified as

$$C_F = C_A(K-1) \quad (3)$$

the factor (K-1) being the increase in the K value over that of air.

Now, the fraction of the total possible fuel quantity in a "linear tank" at any given level is given by L/H, where L is the height of the fuel level and H the total height of the tank. Thus by adding L\H to eqn (3) the complete formula becomes

$$C_F = \frac{L}{H}(k-1)\, C_A \quad (4)$$

The circuit of a basic gauge system is shown in Fig.19.5. It is divided into two sections or loops by a resistance R, both loops being connected to the secondary winding of a power transformer. Loop A contains the tank capacitor $C_T$ and may therefore be considered as the sensing loop of the bridge since it detects current changes due to changes in capacitance. Sensing loop voltage $V_S$ remains constant.

Loop B, which may be considered as the balancing loop of the bridge, contains a reference capacitor $C_R$ of fixed value, and is connected to the transformer via the wiper of a balance potentiometer so that the voltage $V_B$ is variable.

The balance potentiometer is contained within the indicator together with a two-phase motor which drives the potentiometer wiper and indicator pointer. The reference phase of the motor is continuously energized by the power transformer and the control phase is connected to the amplifier and is only energized when an unbalanced condition exists in the bridge.

The amplifier, which may be of either the thermionic-value or the transistor type, has two main stages : one for amplifying the signal produced by bridge imbalance and the other for discriminating the phase of the signal which is then supplied to the motor.

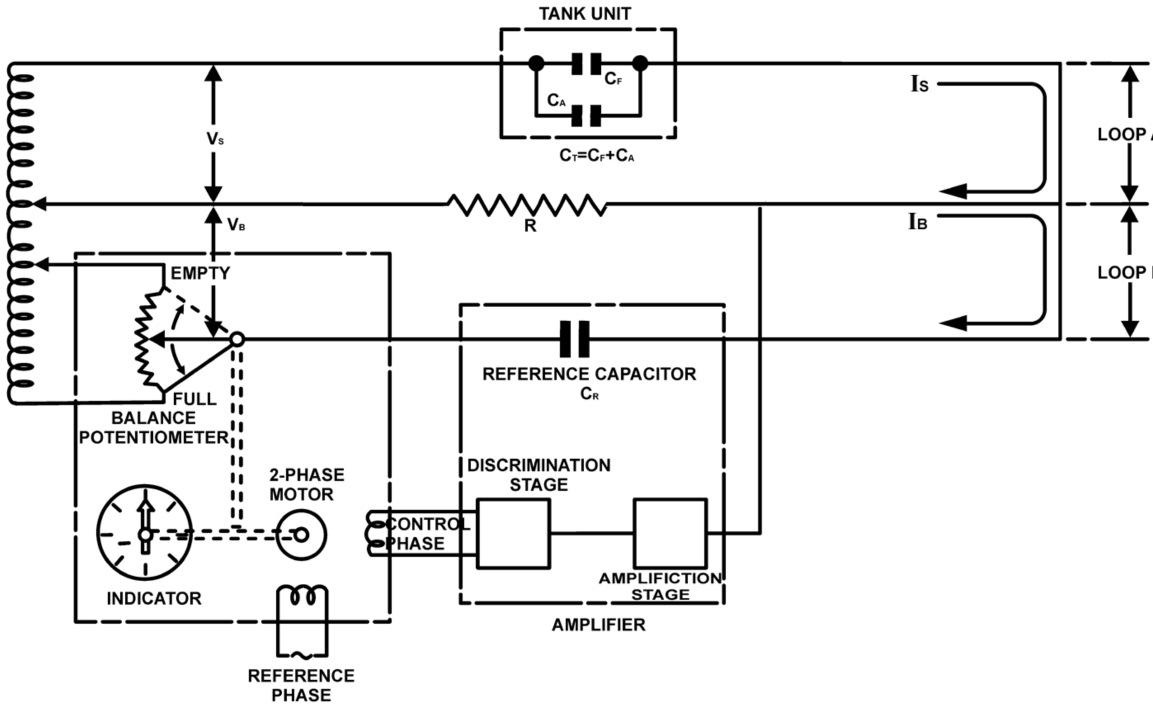



*Fig. 19.5, Circuit of a basic fuel gauge system.*

Let us consider the operation of the complete circuit when fuel is being drawn off from a full tank. Initially, and at the constant full-tank level, the sensing current $I_S$ is equal to the balancing current $I_B$; the bridge is thus in balance and no signal voltage is produced across R.

As the fuel level drops, the tank capacitor has less fuel around it, therefore the added capacitance ($C_F$) has decreased. The tank unit capacitance decreases and so does the sensing current $I_S$, the latter creating an unbalanced bridge condition with balancing current $I_B$ predominating through R.

A signal voltage proportional to $I_BR$ is developed across R and is amplified and its phase detected before being applied to the control phase of the indicator motor. The output signal is a half-wave pulse, a feature of both valves and transistors in discriminatory circuits, and in order to convert it into a full-wave signal, a capacitor is connected in parallel with the control winding. A capacitor is also connected in series with the reference winding to form what is termed a series-resonant circuit. This circuit ensures that the currents in both phase are 90° out of phase, the current in the control phase either leading or lagging the reference phase depending on which loop of the bridge circuit is predominating.

In the condition we are considering the balancing current is predominating : therefore the control-phase current lags behind that of the reference phase causing the motor and balance potentiometer wiper to be driven in such a direction as to decrease the balancing current $I_B$.

When the current $I_B$ equals the current $I_S$, the bridge is once again in balance, the motor stops rotating and the indicator pointer registers the new, lower value.

## PRESSURE INDICATORS

The indicators used for the measurement of pressure in the systems of various type of engine are of two main types : direct-reading and remote indicating. A brief outline of the operating principles of typical indicators and systems is given in the following paragraphs.

### Direct-Reading Indicators

Direct-reading indicators are mechanically operated, and as the name implies, are connected directly to the source of pressure. The most common form of indicator operates on the Bourdon tube principle. A Bourdon tube is a C-shaped tube of oval cross-section having one end closed and the other secured and open to the pressure source. When pressure is admitted to the tube it tend to straighten out causing movement of the closed end and this movement is transmitted to the pointer of the indicator via link and gear mechanism. When applied to low pressure measurements such as manifold pressure and engine power loss, the pressure sensing elements are in the form of flexible bellows or capsules.

**Manifold Pressure Gauges**

Manifold pressure gauges are in most cases of the direct-reading type, designed to measure the absolute pressure at the induction manifold of supercharged engines. The principle of operation and construction of a typical indicator is as follows:

a) The pressure sensing element consists of two metallic bellows mounted in tandem. The rear bellows is connected by a pipeline to the engine induction manifold and the front bellows is evacuated and sealed and is spring-loaded by an internal spring. The outer ends of each bellows are secured to the instrument frame and the inner ends are connected to a common distance piece. The distance piece is connected to a gear type pointer mechanism via an arm, rocker shaft and lever mechanism. Gauges are fitted with a lubber mark which may be pre-adjusted to indicate the maximum manifold pressure permitted for engines with which they are associated.

b) When the engine is stopped and the pressure is at standard conditions, the evacuated bellows assumes a position where its tendency to collapse is balanced by the internal spring. The bellows therefore provides an atmospheric pressure datum against which induction manifold pressure is referenced, The instruments are calibrated in either pounds per square inch, or millimetres or inches of mercury. The zero position on gauges calibrated in the former units corresponds to standard pressure of 14.7 $Ibf/in^2$ while on gauges calibrated in millimetres or inches of mercury, the equivalent value of 760 mm or 29.92 in Hg is read directly. Under engine operating conditions variations in manifold and atmospheric pressures cause relative displacements of the bellows which are transmitted to the pointer via the distance piece, rocker shaft the lever mechanism.

**Torque Pressure Indicators**

These indicators supplement the power indications obtainable from tachometers and manifold pressure gauges by measuring the pressures created by a torquemeter system, such pressures being interpreted as power available at the propeller shaft.

The torquemeter system forms part of the engine itself and is usually built-in with the reduction-gear assembly between the crankshaft and propeller shaft. The construction of a system depends on the type of engine, but in most cases the operation is based on the same principles; i.e. the tendency for some part of the reduction gear to rotate is resisted by pistons working in hydraulic cylinders secured to the gear casing. The principle is shown diagrammatically in Fig. 19.6.

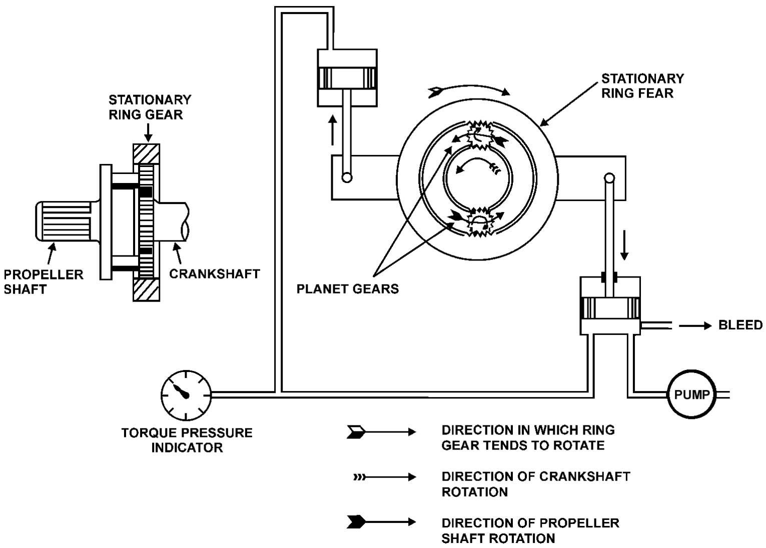



*Fig. 19.6, Principle of torquemeter.*

Oil from the engine system is supplied to the cylinders via a special torquemeter pump and absorbs the loads due to piston movement. The oil is thus subjected to pressures which are proportional to the applied loads or torques and are transmitted to the torque pressure-indicating system, which is normally of the remote-indicating synchronous type.

The brake horsepower is calculated by the following formula :

$$BHP = pN / K$$

Where p is the oil pressure, N the speed (rev./min) and K a torquemeter constant derived from the reduction gear ratio, length of torque arm, and number and area of pistons.

**Power Indicators For Turboprop Engines**

Turboprop engines are, as far as power is concerned, similar to large supercharged piston engines; most of the propulsive force is produced by the propeller, only a very small part being derived from the jet thrust. They are therefore fitted with a torquemeter and pressure gauge system of which the oil pressure readings are an indication of the shaft horsepower. The torquemeter pressure gauge is used in conjunction with the tachometer and turbine gas-temperature indicators.

The indicating system used is governed by the particular type of engine, but there are two main systems in current use and their operating principles are based on the Desynn and alternating-current synchro methods of transmission.

**Desynn Torque Pressure Indicating System**

This system operates on the slab-Desynn principle and is used on Rolls-Royce Dart turboprop engines. In common with the other Desynn systems the transmitter, shown in Fig. 19.7, comprises both mechanical and electrical elements.

The mechanical element consists of a Bourdon tube the open end of which is connected by a flexible hose to the supply line form the oil pump of the engine torquemeter. The free end of the Bourdon tube is connected to the brushes of the electrical element via a sector gear and pinion. A union mounted adjacent to the main pressure connection is connected to a capillary tube accommodated within the Bourdon tube and allows for the bleeding of the system. The transmitter is mounted in a special anti-vibration mounting, the whole assembly being secured to the engine itself.

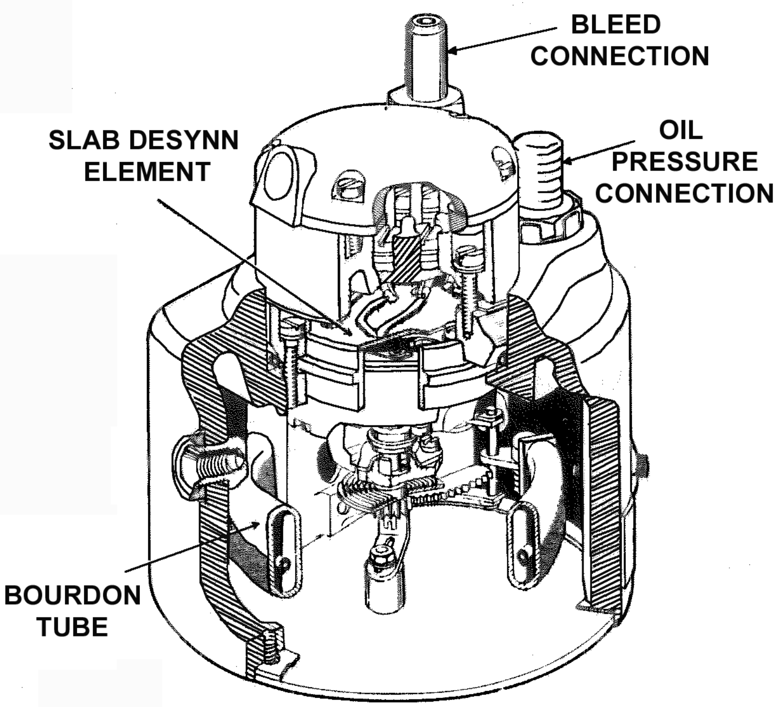



*Fig. 19.7, Desynn torque pressure transmitter.*

With the engine running the pressure produced at the pistons of the torquemeter system is sensed by the transmitter Bourdon tube, the free end of which is distended so as to change the radius of the tube. The movement of the free end is magnified by the sector and pinion, which causes the brushes to be rotated over the slab-wound resistor. The resulting currents, and magnetic field produced in the indicator stator, position the rotor and pointer to indicate the torque pressure on a dial calibrated from 0 to 600 lbf/in$^2$. During operation, and due to pulsations of torquemeter pressure, a certain amount of pointer fluctuation is possible, but this is limited to 30 lbf/in$^2$ on either side of a mean torquemeter pressure reading.

**Synchro Torque Pressure Indicating System**

This system is an application of the alternating-current control synchro principle.

As may be seen from Fig. 19.8, the mechanical element of a synchro torque pressure transmitter is very similar to that of the slab-Desynn type. The Bourdon tube, sector gear and pinion, however, are arranged to drive the rotor of a CX synchro. The transmitter is designed for mounting directly on to an engine and is connected by flexible tubing to the torquemeter system.

The indicator consists of a CT synchro connected to the transmitter synchro CX, a two-stage transistor amplifier, a two-phase servomotor, and two concentrically mounted pointers driven through a gearbox. The smaller pointer indicates hundreds of pounds and rotates in step with the synchro rotor, while the larger pointer rotates ten times as fast.

When the Bourdon tube senses a change in torquemeter system pressure it causes the CX rotor to rotate and to induce a signal voltage in its stator which is then transmitted to the stator of the CT synchro in the indicator. This signal results in a change in direction of the resultant magnetic field with respect to the CT rotor position, thus inducing an error voltage signal in the rotor. The error signal is fed to the amplifier, which determines its direction, i.e. whether it results form an increase or a decrease in pressure, as well as amplifying it. The amplified signal is then applied to the control phase of the servomotor, which, via the gearbox, drives the pointers in the appropriate direction also drives the CT rotor until it reaches a new null position at which no further error voltage signal is induced.

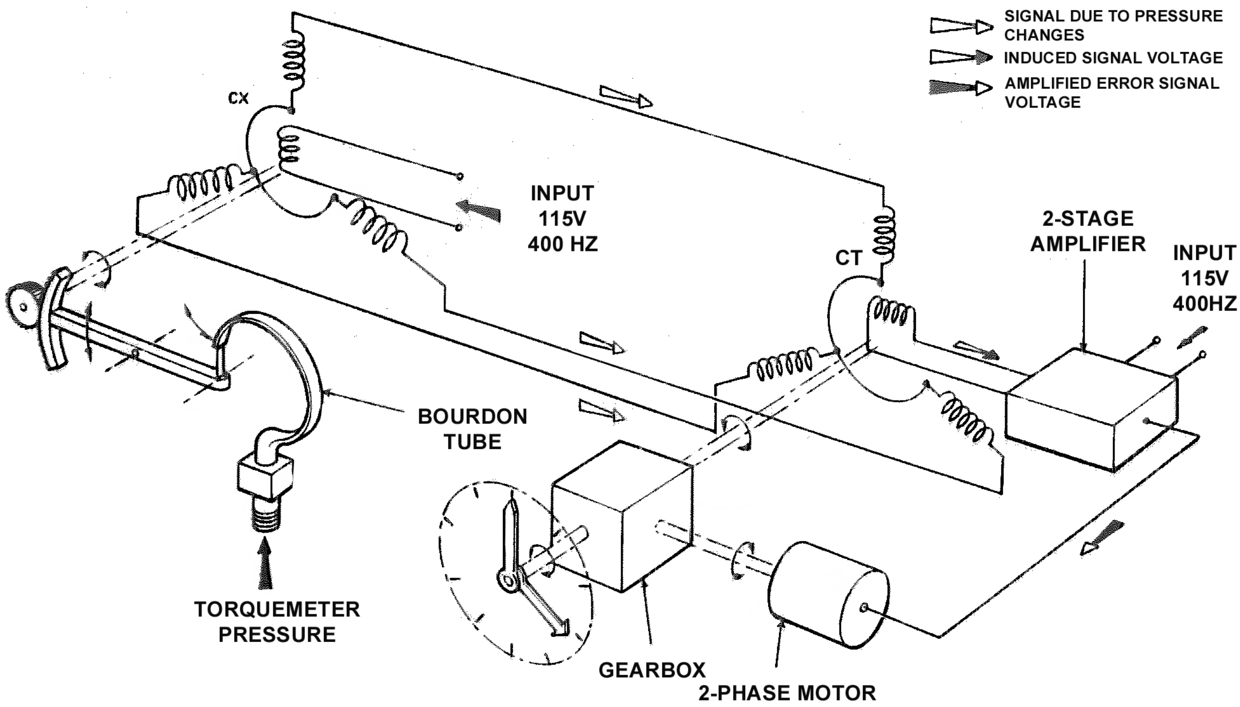



*Fig. 19.8, Principle of control synchro-type torque pressure indicating system.*

**Power Loss, Engine Pressure Ratio and Percentage Thrust Indicators**

These indicators are sensitive differential pressure gauges designed to indicate the thrust output of turbojet engines in terms of the absolute pressure within the exhaust unit. The application of each type of indicator depends on the type of engine and installation. A typical indicator consists of two capsules connected by a lever and gear train assembly to a pointer which rotates over a scale calibrated in either in. of Hg absolute (Power Loss and Engine Pressure Ratio Indicators) or percentage of thrust. One of the capsules is connected by a sensing line to the exhaust unit and responds to the difference between exhaust unit pressure and static pressure supplied to the indicator case. The other capsule is of the aneroid type and is sensitive to static pressure only which, in the case of power loss indicators and percentage thrust indicators, is supplied from a static vent of the pitot-static system. In the case of engine pressure ratio indicators static pressure is supplied from a vent at the engine air intake. The deflection of the lever system in response to displacements of both capsules is equal to exhaust unit pressure, and is transmitted to the pointer via the gear train.

**Engine Gauge Unit**

The engine gauge unit is comprised of three separate instruments housed in a single case. A typical engine gauge unit, containing gauges for oil and fuel pressure and oil temperature, is shown in figure 19.9.

Two types of oil temperature gauges are available for use in an engine gauge unit. One type consists of an electrical resistance type oil thermometer, supplied electrical current by the aircraft d.c. power system. The other type, a capillary oil thermometer, is a vapour pressure type thermometer consisting of a bulb connected by a capillary tube to a Bourdon tube. A pointer, connected to the Bourdon tube through a multiplying mechanism, indicates on a dial the temperature of the oil.

The Bourdon tube is an aircraft instrument made of metal tubing, oval or somewhat flattened in cross section (figure 19.10). The metal tubing is closed at one end and mounted rigidly in the instrument case at its other end.

The fluid whose pressure is to be measured is introduced into the fixed end of the Bourdon tube by a small tube leading from the fluid system to the instrument. The greater the pressure of the fluid, the more the Bourdon tube tends to become straight. When the pressure is reduced or removed, the inherent springiness of the metal tube causes it to curve back to its normal shape.

If an indicator needle or pointer is attached to the free end of the Bourdon tube, its reactions to changes in the fluid pressure can be observed.

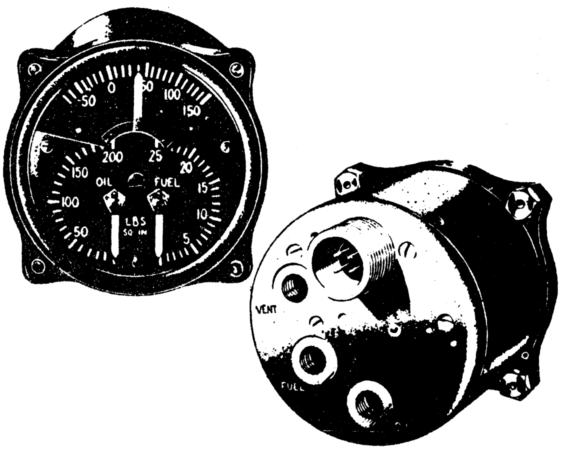


*Fig. 19.9. Engine gauge unit.*

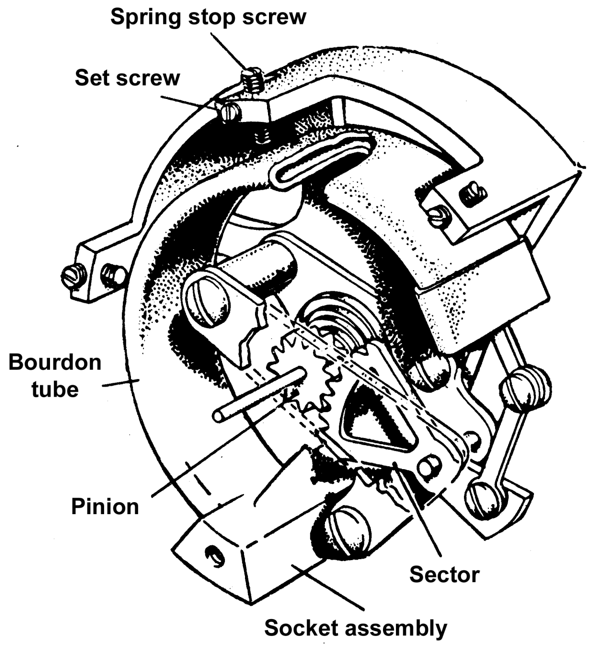



*Fig. 19.10, Bourdon tube pressure gauge*

**Hydraulic Pressure Gauge**

The mechanisms used in raising and lowering the landing gear or flaps in most aircraft are operated by a hydraulic system. A pressure gauge to measure the differential pressure in the hydraulic system indicates how this system is functioning. Hydraulic pressure gauges are designed to indicate either the pressure of the complete system or the pressure of an individual unit in the system.

A typical hydraulic gauge is shown in figure 19.11. The case of this gauge contains a Bourdon tube and a gear-and-pinion mechanism by which the Bourdon tube's motion is amplified and transferred to the pointer. The position of the pointer on the calibrated dial indicates the hydraulic pressure in p.s.i.

The pumps which supply pressure for the operation of an aircraft's hydraulic units are driven either by the aircraft's engine or by an electric motor, or both. Some installations use a pressure accumulator to maintain a reserve of fluid under pressure at all times. In such cases the pressure gauge registers continuously. With other installations, operating pressure is built up only when needed, and pressure registers on the gauge only during these periods.

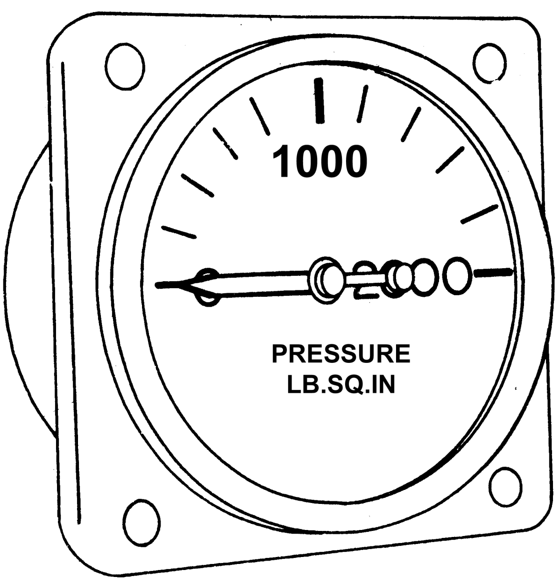



*Fig. 19.11, Hydraulic Pressure gauge.*

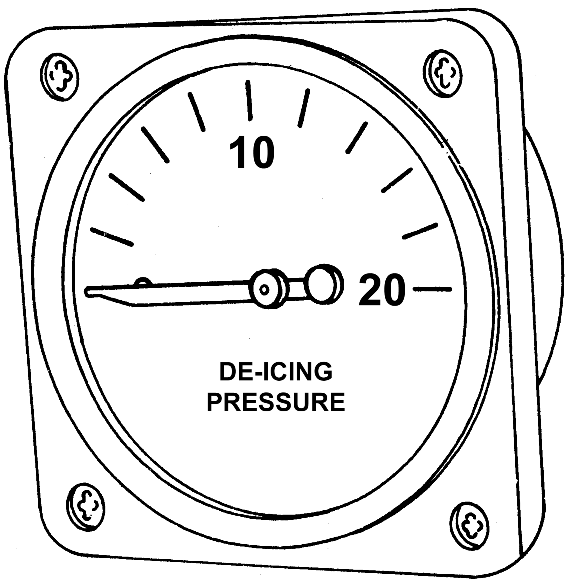



*Fig. 19.12, De-icing pressure gauge.*

**De-icing Pressure Gauge**

The rubber expansion boots, which de-ice the leading edges of wings and stabilizers on some aircraft, are operated by a compressed air system. The de-icing system pressure gauge measures the difference between prevailing atmospheric pressure and the pressure inside the de-icing system, indicating whether there is sufficient pressure to operate the di-icer boots. The gauge also provides a method of measurement when adjusting the relief-valve and the regulator of the de-icing system.

A typical de-icing pressure gauge is shown in figure 19.12. The case is vented at the bottom to keep the interior at atmospheric pressure, as well as to provide a drain for any moisture which might accumulate.

The pressure-measuring mechanism of the de-icing pressure gauge consists of a Bourdon tube and a sector gear, with a pinion for amplifying the motion of the tube and transferring it to the pointer. The de-icing system pressure enters the Bourdon tube through a connection at the back of the case.

The range of the gauge is typically from zero p.s.i. to 20 p.s.i., with the scale marked in 2-p.s.i. graduations as shown in figure 19.12.

When installed and connected into an aircraft's de-icing pressure system, the gauge reading always remains at zero unless the di-icing system is operating. The gauge pointer will fluctuate from zero p.s.i. to approximately 8 p.s.i. under normal conditions, because the de-icer boots are periodically inflated and deflated. This normal fluctuation should not be confused with oscillation.

**REMOTE-INDICATING INSTRUMENTS**

These instruments measure pressure at source and transmit the measured values to panel-mounted instruments through the medium of either a liquid or electrical signal transmission system. Depending on the type of aircraft the liquid transmission system may either be of the closed capillary type, or pressure transmitter type. Remote indicating instruments employing the electrical signal transmission principle consist of separate transmitting and indicating units which may form part of a d.c. or a.c. synchronous data transmission system, or a ratiometer system. The pressure sensing elements of a transmitter may either be a flexible metal bellows or a Bourdon tube, depending on the application of the indicating system.

**a) Capillary Type**

The capillary system consists of a transmitter unit containing a capsule which is connected by a length of capillary tubing to a Bourdon tube indicator mechanism. The capsule, capillary tubing and Bourdon type are completely filled with a special fluid such as Heptane (a paraffin hydrocarbon having a low freezing point and very little viscosity change) the whole assembly forming a single sealed system. The transmitter unit is directly connected by means of a hollow bolt to the particular engine system (e.g. oil system) and when pressure is admitted a force is exerted on the capsule to cause displacement of the transmitting fluid. This displacement in turn tends to straighten out the Bourdon tube in the manner of a direct-reading pressure gauge.

**b) Pressure Transmitter System**

A pressure transmitter system functions in a similar manner to the capillary system, but has the advantage that the transmitting and indication units are separate thus facilitating removal and installation and also permitting filling of the system in situ. The transmitter unit consists of two flanged circular housing bolted together and divided into an inlet chamber and an outlet chamber by a neoprene diaphragm. The inlet chamber is connected to the pressure source, while the outlet chamber is connected via small bore tubing to a direct-reading Bourdon tube pressure gauge. A spring-loaded ball valve is teed into the gauge connection and serves as a bleed during filling operations. The outlet chamber, tubing and Bourdon tube are filled with a mineral base oil. A second spring-loaded ball valve is incorporated in a connection in the lower part of the outlet chamber, the connection being used for filling purposes. A metal centralising disc in the outlet chamber prevents distension of the diaphragm when the system is being filled, and may be repositioned by a centralising knob.

i) In operation, pressure is exerted on the diaphragm causing it to distend and to transmit the pressure to the fluid in the outlet chamber. The fluid tends to force itself out of the chamber, but as the system is closed one the Bourdon tube is displaced and the indicator reads the pressure applied at the transmitter.

**c) Ratiometer pressure Indicators**

These instruments are used principally for the measurement of pressures in engine fuel and oil systems, and depending on the type specified, operate from either a direct current or alternating current source. In both cases the ratio of currents flowing through two coils of the indicator is measured in terms of pressure. The transmitter units normally consist of a pressure sensing bellows assembly mechanically coupled to an electrical element which in the case of a direct current system takes the form of a modified Micro-Desynn element connected to a moving coil ratiometer. The electrical element of an alternating current operated transmitter consists of inductor coils the iron cores of which are positioned by the bellows assembly. The indicator consists of two coils connected to the transmitter coils and two cam-shaped discs free to rotate in the air gaps of the cores. The discs are mounted on a common shaft connected to the pointer. When pressure at the transmitter changes it causes an increase in current in one coil circuit and a decrease in the other, the discs rotating in a direction determined by the coil carrying the increased current. Rotation of the discs and pointer ceases when a balance is reached by the torques produced at the discs.

**d) Pressure Switches**

Pressure switches are used in conjunction with warning lights to indicate that the pressure in a particular fluid system (e.g. fuel, oil and water methanol systems) has reached a pre-determined value. Depending on the application this value may be the upper or lower limit of the safe working pressure. Some switches are also employed to initiate a sequence of controlling operations when a certain pressure is reached in a fluid system.

i) A typical unit consists of a metal diaphragm sandwiched between the flanges of a contact box which forms the main body and a base through which the system fluid is admitted, and a fixed and moving contact switch assembly. Connection of the base to the pressure source varies but, in general , it is either of the banjo fitting type or direct flange-mounted type. The movable contact to actuated by a push rod bearing against the upper surface of the diaphragm. The stationary contact is adjustable to provide various pressure settings. Access to the adjusting screw is gained by removal of an access plate in the cover of the contact box, or as in some units by removal of the cover itself.

ii) The pressure of the fluid entering the base of the switch unit acts against the diaphragm causing the push-rod to operate the movable contact. The contact position is thus changed to either make or break the warning or controlling circuit.

**ENGINE SPEED INDICATORS**

These indicators measure the rotational speeds of piston engine crankshafts and compressor shafts of turbine engines. Two principal types are in use; mechanical as used in some types of single engine aircraft, and electrical which are used in aircraft having multi-arrangements of piston engines, and in all turbine-powered aircraft.

**Mechanical Indicators**

Mechanical indicators consist of a flyweight assembly connected to the engine by a flexible drive shaft and coupled to a gear type pointer mechanism. The gear ratios at the engine and indicator are such that the flexible drive shaft rotates at a lower speed to minimise wear. As the shaft rotates centrifugal forces act on the flyweight and cause it to take up a certain angular position. The displacement is transmitted to the pointer which rotates over the scale to indicate the speed of the engine crankshaft.

**Flexible Drive Shafts**

A flexible drive consists of a flexible inner shaft which is free to rotate within a stationary outer casing. The inner shaft embodies a central core of hardened steel wire over which five layers each of four strands of finer gauge wire are wound, alternate layers being wound in opposite direction. After the inner shaft has been cut to the appropriate length a connector is squared to each end by swaging. Both connectors incorporate squared shanks which engauge the hollow squared ends of the engine drive shaft and indicator shaft respectively. The shank at the engine end of the drive is longer than that at the indicator end.

**Note**- In some types of flexible drive, the connector shank at the engine end is designed to engauge with a keyway in the engine drive shaft, while the shank at the indicator end is hollow.

The outer casing is a continuous winding of two specially formed steel wires and is flexible, oil tight and waterproof. Flanged collars are swaged to each end of the outer casing to provide a means of attachment to the engine and indicator. Axial movement of the inner shaft is restricted by a shoulder on the connector at the indicator end which abuts on the end of the flanged collar and also by an interposed slip washer. This arrangement permits considerable end float of the shaft in the outer casing.

**Electrical Indication Systems**

Electrical indicating systems comprise an alternating current generator which supplies a synchronous motor-driven indicator. The generator consists of a permanent magnet rotor and a three-phase stator winding. The rotor may be driven by a short length flexible drive shaft, or as in the case of turbine engines which have high rotational speeds prohibiting the use of flexible drives, by direct coupling to a splined shaft driven by the compressor shaft via reduction gearing.

The synchronous motor of an indicator is coupled to an eddy current drag type of mechanism consisting of a permanent magnet, a cup shaped or disc type of drag element, and a controlling spring. The drag element is mounted on a spindle connected to a gear mechanism which drives a large and a small pointer to indicate hundreds and thousands of revolutions per minute respectively.

Rotation of the generator rotor induces a three-phase voltage in the stator windings which is transmitted to the winding of the indicator synchronous motor causing the rotor to revolve at a speed proportional to the generator frequency and therefore engine speed. The permanent magnet of the drag mechanism is also rotated and induces eddy currents in the drag element tending to rotate it at the same speed as the magnet. As the controlling spring is coupled to the drag element spindle it restrains rotation of the element to a position at which spring force and drag torque are in balance. The pointers are therefore positioned to indicate the engine speed.

**Percentage Speed Indicators**

These indicators are designed to indicate the speed of a turbine engine as a percentage of the nominal maximum speed. The scales are graduated from 0 to 100 % , the 100% indication corresponding to the nominal maximum engine speed and a specific generator drive speed which is usually 4,200 rev/min. Indicators allow for slight increases in nominal maximum engine speed by reading up to 110%. In principle they operate in a similar manner to the conventional alternating current type of indicating system.

**Synchroscopes**

Synchroscopes are designed for use in multi-engined aircraft to indicate the degree of synchronism existing between a selected 'master' engine and the remaining engines designated as 'slaves'. They form part of an engine speed indicating system, each engine being associated with a complete synchroscope unit housed within the instrument.

A typical unit consists of a synchronous motor having a three-phase starwound stator and rotor. A small double-ended pointer is attached to the rotor shaft and is referenced against a dial markedSLOW to the left and FAST to the right. In some types of synchroscope the left and the right position of the dial are marked INCREASE and DECREASE respectively.

The generator of the selected 'master' engine is electrically connected to the synchroscope rotor while the 'slave' engine generator is connected to the stator. The output from the master engine generator induces a rotating magnetic field in the synchroscope rotor at a frequency proportional to the generator speed. Similarly, the generator of the slave engine induces a rotating magnetic field in the synchroscope stator. Both fields rotate in the same direction, and under synchronised speed conditions they interact to maintain the rotor and pointer at some stationary position. When there is a difference between generator speeds field interaction causes the rotor to rotate at a speed equal to the difference, and in a clockwise or anti-clockwise direction according to whether the speed of the slave engine is greater or less than the master engine speed.

**Rotation Indicators**

In turbine engines of the by-pass type severe damage may arise if the low-pressure shaft is not free to rotate during the starting cycle; the damage being caused by the re-circulation of hot gases around the high-pressure system. In order to indicate that the low-pressure shaft begun to rotate and that it is safe to continue the starting cycle, rotation indicators are provided.

The basis of an indicator is an amplifier connected to one phase of the engine speed indicator system to accept signals from it as a speed reference input. The output stage is connected to an indicator lamp located on the main instrument panel, or on a panel at a flight engineer's station. Depending on the design, an amplifier may require a power supply of 115-volts at 400 Hz, or may be completely independent of aircraft power supplies.

When the input reaches a critical level, the amplifier produces sufficient output to light the indicator lamp. A typical critical input level is 6 mv, corresponding to a rotation speed of a fraction of 1 rev/min. This speed is reached in the first few degrees of rotation and the lamp starts flashing immediately the shaft begins to rotate. Input signals in excess of the critical causes the amplifier to saturate and the lamp to remain alight but without being overloaded.
An indicator is only required during the starting cycle and for this reason the power supply to the amplifier is fed via an engine starting circuit. For multi engine installations, a single amplifier and indicator lamp serves to indicate rotation of each engine which is automatically selected during each starting cycle.

**TEMPERATURE INDICATORS**

All temperature indicators are of the transmitting type and fall into three main categories:

a) Capillary.
b) Electrical Resistance.
c) Thermo-Electric.

**Capillary Thermometers**

Capillary thermometers are used to measure the temperature of liquids in aircraft systems such as oil and coolant and although superseded by electrical resistance, thermometers are still fitted to certain types of small aircraft. In general the construction is similar to a liquid transmission type of pressure indicator which consists of a transmitter unit joined to a Bourdon tube indicator by a length of capillary tubing. In this case the transmitter unit is in the form of a 'bulb' which is immersed in the fluid whose temperature is to be measured. The transmitting fluid contained within the system being either mercury or ethyl ether. When the bulb temperature changes the mercury expands or contracts, or in the case of ethyl ether the vapour pressure increases or decreases, causing displacement of the Bourdon tube and corresponding movement of the pointer.

**MERCURY-IN-STEEL THERMOMETER**

The mercury-in-glass thermometer is not practicable for use in aircraft although it is used as a standard against which aircraft thermometers are checked. A more practical temperature-measuring instrument is the mercury-in-steel thermometer. This comprises a steel bulb, capillary tube and Bourdon tube completely filled with mercury. A change of temperature causes a change of volume of mercury and the Bourdon tube moves to show the change. A pointer moves over a scale to indicate the temperature, usually in deg. C. in the case of aircraft thermometers (Fig. 19.13.).

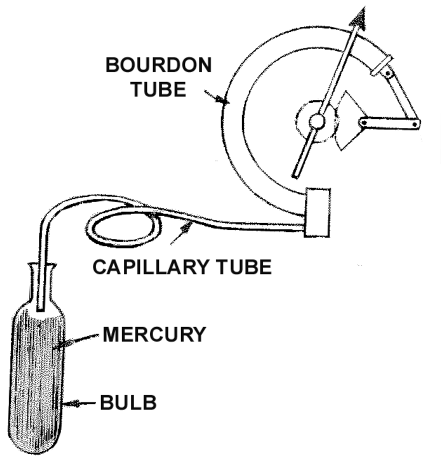



*Fig. 19.13.*

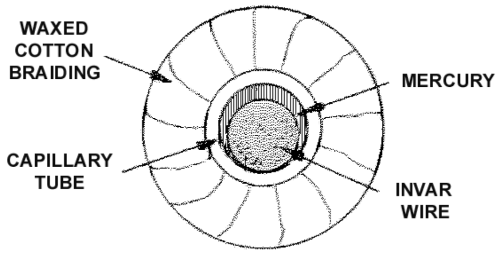



*Fig. 19.14.*

***Compensation for Errors***

Errors in this instrument arise mainly from the fact that the mercury in the capillary tube and indicator is also sensitive to temperature changes and these will be registered on the dial. To reduce the errors to a minimum, the volume of mercury in the capillary tube and Bourdon tube is kept small in comparison with the volume of the bulb. Compensation for remaining errors is effected in several ways, a common method being as follows. Compensation for temperature errors in the capillary tube can be achieved by running a fine Invar wire down the length of the capillary so that the cross-section of the tube would be as in Fig. 19.14. The effect of this is that an increase of temperature of he mercury in the capillary tube causes an increase of volume which would normally show as an increased reading on the indicator. However, because of the differential expansion of the steel tube and Invar wire (the expansion of which remains practically constant), an increase in the volume of the space between the wire and the walls of the tube occurs and this accommodates the increase in volume of mercury. Compensation in the indicator can be arranged by fitting a bi-metal strip to the mechanism which counteracts the effect of changes of volume of mercury in the indicator.

***Double-Spiral Bourdon Tube***

In the double-spiral Bourdon-tube instrument there is no mechanism, and compensation in his case takes the form of a bi-metal helix which forms the last coil of the double-spiral tube (Fig. 19.15).

There is little maintenance that can be done on this instrument, and because of its construction as a system comprising bulb, capillary tube and Bourdon tube, it is difficult to remove from the aircraft. Calibration checks, can, however, be carried out by immersing the bulb in an oil bath together with standard mercury-in-glass thermometer. Readings are compared as the oil bath is slowly heated, care being taken to agitate the bath so as to obtain a uniform temperature before taking any readings. The errors of the instrument are in the region of 2 percent.

The main use of the mercury-in-steel thermometer is as an oil-temperature gauge indicating the temperature of the engine lubricating oil. It is also used as an air-temperature gauge for indicating the temperature of air outside the aircraft, so that the navigator has a means of applying corrections to airspeed, altitude and so on. On larger modern aircraft these mechanical temperature gauges have been superseded by the electrical types of thermometer which have the advantage of easier maintenance because the various units can be disconnected from a terminal block.

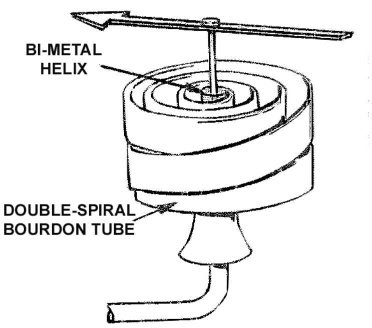



*Fig. 19.15.*

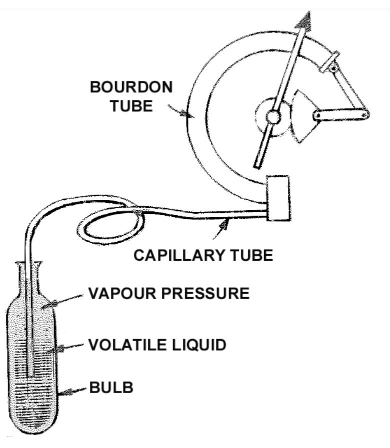



*Fig. 19.16.*

**VAPOUR-PRESSURE THERMOMETER**

Another mechanical temperature gauge, or thermometer, is the vapour-pressure thermometer. This comprises a thermometer bulb, capillary tube and Bourdon tube as in the mercury system, but the operating medium is a highly volatile liquid such as ethyl ether. The volume of liquid contained in the system is not sufficient completely to fill it and the remaining space contains vapour given off by the chosen liquid ( Fig. 19.16).

A rise in the temperature of the thermometer bulb increases the evaporation rate of the liquid to produce a greater vapour pressure. This pressure is transmitted, via the liquid, to the Bourdon tube, thus indicating the value of pressure. Since the pressure is proportional to the bulb temperature, the indicator can be calibrated in terms of Fahrenheit or Celsius degrees. The scale range of this instrument is limited by the fact that the vapour pressure is negligible at normal temperatures, and a common range for this instrument is +40 to +140 deg. C. In this form the instrument is often used as a coolant thermometer, the bulb being immersed in the radiator header tank.

The instrument is calibrated in the same manner as the mercury-in-steel thermometer; i.e. the bulb is immersed in a tank containing a master mercury-in-glass thermometer. The tank is heated and the readings of the vapour-pressure gauge are compared with the readings of the mercury thermometer. Maintenance is the same as for normal pressure gauges, although care should be taken in handling the system, particularly in respect of the capillary tube. The slightest "crack" in the system, or a system where the capillary tube is not properly soldered into the bulb and indicator, will allow evaporation of the fluid. This results in a larger space being left to be filled by the vapour and consequently a lower pressure indication for a given temperature. This is not an unusual fault with this type of instrument.

In the mercury type of thermometer it was shown that the system comprising bulb, capillary tube and Bourdon tube was made of steel. This is because mercury forms an amalgam with most other metals. There is no necessity for such a system in the case of the vapour-pressure thermometer and the bulb is usually of brass, while the capillary is of annealed copper; the Bourdon tube may be phosphor bronze or beryllium copper.

**Electrical Resistance Thermometers**

Thermometers of this type are comprised of separate bulb and moving coil indicator electrically interconnected and supplied with direct current from the aircraft electrical system. The bulb contains a coil of nickel or platinum wire which forms a variable resistance arm of a bridge circuit contained within the indicator. When the bulb temperature changes, the coil resistance increases or decreases causing current to flow through the moving coil system corresponding to the temperature resistance characteristics of the bulb material used. The circuit arrangements may be either of the Wheatstone Bridge or ratiometer type.

***Variable Resistance systems***

A system consists of a sensor unit (generally referred to as a 'bulb') and an indicator, connected in a series circuit configuration, and requiring dc power which may be directly supplied from a relevant busbar or, in some cases, by rectification of a single-phase ac supply. Sensor units employ resistance coils of either nickel or platinum wire, and the indicator units are of the moving coil type, having their internal circuits arranged in either the basic Wheatstone bridge configuration, or in the more commonly adopted ratiometer configuration.

***Sensor Units***

the general arrangement of a sensor unit commonly used for the measurement of liquid temperatures is shown schematically in Fig. 19.17. The resistance coil is wound on an insulated former and the ends of the coil are connected to a two-pin socket (or a plug) via contact strips. The 'bulb', which serves as a casing to protect and to seal the coil assembly, is a stainless steel tube closed at one end and secured to a union nut at the other. The union nut is used for securing the complete unit to the connecting point at which the temperature is to be measured.

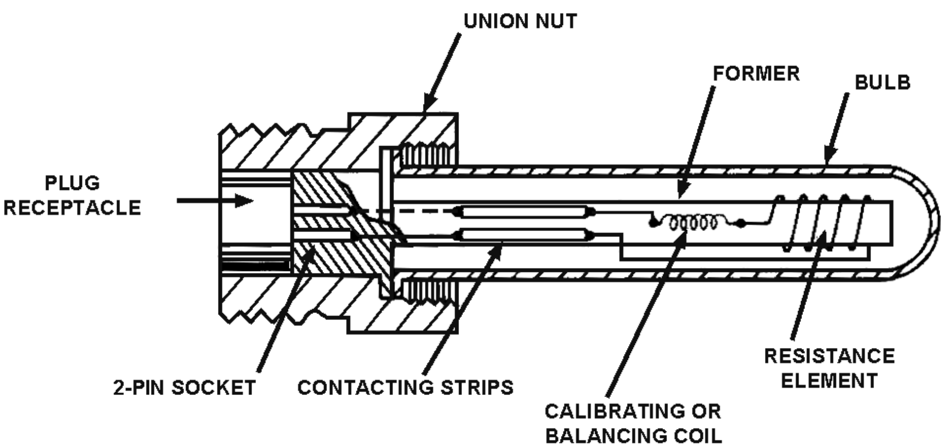



*Fig. 19.17, Schematic arrangement of a temperature sensor.*

It will be noted from the diagram that the coil is wound at the bottom end of its former and not along the full length. This ensures that the coil is well immersed in the hottest part of the liquid, thus minimizing errors due to radiation and conduction losses in the 'bulb'.

A calibrating or balancing coil is normally provided so that a standard constant temperature/resistance characteristic can be obtained, thus permitting interchangeability of sensor material. The coil, which may be made from Manganin or Eureka wire, is connected in series with the sensor coil and is pre-set in value during initial calibration by the manufacturer.

**Wheatstone Bridge Circuit**

In this arrangement a measure of the temperature at various points throughout the range, and for a given supply voltage, is obtained in terms of an out-of-balance current. By suitable arrangement of the bulb material the bridge may be balanced at a predetermined temperature and no current will flow through the indicator. This is known as the 'null point' and in general is used to indicate the critical temperature of the instrument since, when these conditions prevail, the indication is independent of supply voltage. At all other points on the scale the out-of balance current depends not only on the bulb resistance but also on supply voltage; therefore, there will be an error in indicated reading if this voltage differs from that for which the instrument was calibrated. This error will be proportional to the percentage difference change in the supply voltage from that used in calibration ,and to the amount by which the pointer is deflected from the 'null point' A device for adjusting the moving coil and pointer to the 'null point' is always provided and may be set by a screw at the front of the indicator bezel.

The Wheatstone-bridge meter operates on the principle of balancing one unknown resistor against other known resistance. A simplified form of a Wheatstone-bridge circuit is shown in figure 19.18. Three equal values of resistances (A, B and C, figure 19.18) are connected to a diamond-shaped bridge circuit with a resistance of unknown value (D).

The unknown resistance represents the resistance of the temperature bulb of the electrical resistance thermometer system. A galvanometer calibrated to read in degrees is attached across the circuit at point X and Y.

When the temperature causes the resistance of the bulb to equal that of the other resistances, no potential difference exists between points X and Y in the circuit, and no current flows in the galvanometer leg of the circuit. If the temperature of the bulb changes, its resistance will also change, and the bridge becomes unbalanced, causing current to flow through the galvanometer in one direction or the other.

The dial of the galvanometer is calibrated in degrees of temperature, converting it to a temperature-measuring instrument. Most indicators are provided with a zero adjustment screw on the face of the instrument to set the pointer at a balance point (the position of the pointer when the bridge is balanced and no current flows through the meter).

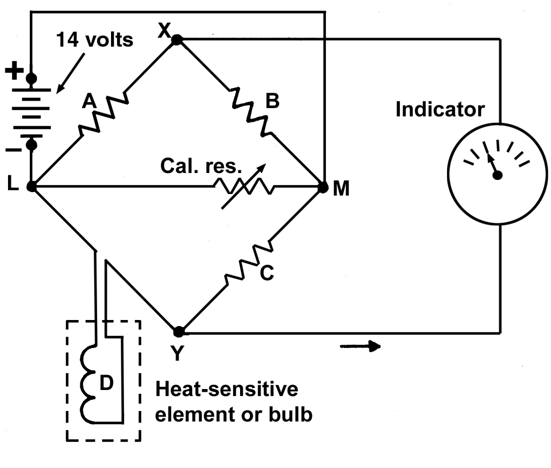



*Fig. 19.18. Wheatstone-bridge meter circuit.*

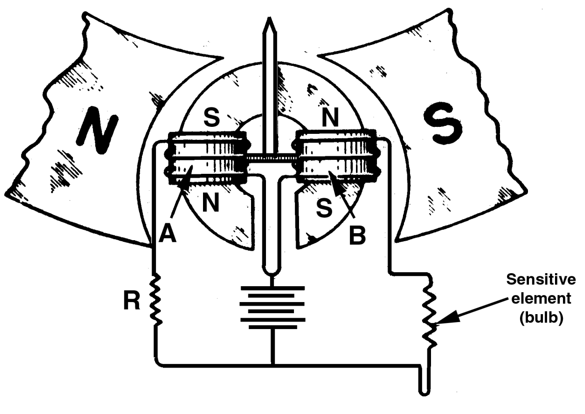



*Fig. 19.19. Ratiometer temperature-measuring system schematic.*

**Ratiometer Circuit**

In the ratiometer circuit arrangement the moving coil system is made up of two coils rotating in a magnetic field which, unlike conventional moving coil instruments, is non-uniform. One of the coils carries a reference current while the other is connected to the resistance bulb and therefore carries a current proportional to the resistance and temperature of the bulb. The ratio of these two currents determines the position of the complete coil assembly and the pointer, the indication being virtually independent of variations in the supply voltage.

**RATIOMETER ELECTRICAL RESISTANCE THERMOMETER**

The basic Wheatstone-bridge, temperature -indicating system provides accurate indications when the pointer is at the balance point on the indicator dial. As the pointer moves away from the balance point, the Wheatstone-bridge indicator is increasingly affected by supply voltage variations. Greater accuracy can be obtained by inserting one of several types of automatic line voltage compensating circuits into the circuit. Some of these voltage regulators employ the filament resistance of lamps to achieve a more uniform supply voltage. The resistance of the lamp filaments helps regulate the voltage applied to the Wheatstone-bridge circuit since the filament resistance changes in step with supply voltage variation.

The ratiometer is a more sophisticated arrangement for obtaining greater accuracy in resistance-bulb indicators. The ratiometer measures the ratio of currents, using an adaptation of the basic Wheatstone-bridge with ratio circuitry for increased sensitivity.

A schematic of a ratiometer temperature circuit is shown in figure 19.19. The circuit contains two parallel branches, one with a fixed resistance in series with coil A, and the other a built-in resistance in series with coil B. The two coils are wound on a rotor pivoted in the centre of the magnet air gap. The permanent magnet is arranged to provide a larger air gap between the magnet and the coils at the bottom than at the top. This produces a flux density that is progressively stronger from the bottom of the air gap to the top.

The direction of the current through each coil in respect to the polarity of the permanent magnet causes the coil with the greater current flow to react in the weaker magnetic field. If the resistance of the temperature bulb is equal to the value of the fixed resistance, and equal values of current are flowing through the coils, the torque on the coils will be the same and the indicator points will be in the vertical (zero) position.

If the bulb temperature increases, its resistance will also increase, causing the current through the coil B circuit branch to decrease. Consequently, the torque on coil B decreases and coil A pushes downward into a weaker magnetic field; coil A, with its weaker current flow, moves into a stronger magnetic field. The torques on the coils still balance since the product of current times flux remains the same for both coils, but the pointer has moved to a new position on the calibrated scale. Just the opposite of this action would take place if the temperature of the heat-sensitive bulb should decrease.

Ratiometer temperature-measuring systems are used to measure engine oil, outside air, and carburettor air temperatures in many types of aircraft. They are especially in demand to measure temperature conditions where accuracy is important of large variations of supply voltages are encountered.

**Thermoelectric Systems**

Thermoelectric systems are utilised principally for the measurement of air-cooled piston engine cylinder head temperatures and exhaust gas temperatures of turbine engines. The system comprises a single or multiple thermocouple arrangement located at the appropriate source of temperature, a moving coil millivoltmeter calibrated to relevant temperature/e.m.f. characteristics in degrees C, and connecting cables of known length and resistance.

**Principle**

A thermocouple assembly is made up of two dissimilar metal conductors joined together to form a hot junction. The open ends of the conductors are connected by cables to the indicator which forms a cold junction. When the hot junction is subjected to temperature an emf is generated which causes a current to flow in the closed circuit. The magnitude of the e.m.f. depends on the materials used for the thermocouple and the difference between the hot and cold junction temperature.

**Cold-Junction Temperature Compensation**

To avoid errors in indicator readings due to the effects of temperature changes at the cold junction, automatic compensation devices are fitted to thermo-electric system indicators. Three methods are normally employed,
(i) Mechanical, (ii) Electrical and (iii) Magnetic; methods (ii) and (iii) Compensating for changes in moving coil resistance to ensure proper instrument current.

(a) The mechanical method directly compensates for cold-junction temperature changes and consists of a bimetallic spiral, the outer end of which is connected to the outer end of the controlling hairspring. When the cold-junction temperature changes the magnitude of the circuit e.m.f. changes and causes an error in the required hot-junction temperature indications. For example, if it increases, the difference between hot-and cold-junction temperatures, and circuit e.m.f, is reduced and the moving coil and pointer tends to indicate a lower hot-junction temperature. The bimetallic spiral is also affected by the cold-junction temperature change, but as it is wound in a direction opposite to that of the controlling hairspring it opposes the moving coil and a constant hot-junction temperature indication is maintained.

(b) In the electrical method a neutraliser coil is connected in series with the moving coil, the characteristics of the material being such that under the same temperature conditions its resistance change equally opposes that of the moving coil material. In some instruments a thermistor shunted by a Eureka coil serves as the compensator.

(c) The magnetic method of compensation is accomplished by means of temperature sensitive magnetic strips (magnetic shunt) clamped across the permanent magnet of the indicator. These have the effect of shunting the magnet poles so that the flux strength in the air gap is varied at a rate proportional to the rate at which moving coil resistance changes.

**Engine Cylinder Head Temperature Indicating Systems**

The metal combinations used for cylinder head temperature thermocouples are either copper/constantan or iron/constantan. Depending on the type of engine the hot-junction may be formed either for bolting under a sparking plug or in direct contact with a cylinder head. A thermocouple is attached to the cylinder, which tests have shown to be the hottest on an engine in any particular installation. The indicators are of the semicircular scale type usually calibrated over the range 0-350°C, and are provided with terminal type connections at the rear of their cases. The terminal identified by a positive sign is connected to a copper or iron lead depending on the thermocouple combination used. Adjustment of the pointers to prevailing cold-junction temperatures is affected by a device located at the front of the instrument and which is adjusted by a screwdriver.

**Turbine Exhaust Gas Temperature Indicators**

Gas temperature is a critical variable of turbine engine operation and it is essential to provide an indication of this temperature. The two control positions in an engine at which measurement are normally taken are: (i) At the exhaust unit (ii) Within the turbine at one of the stator positions.

(a) Several factors have to be considered before any position is adopted; for example, temperatures related to the performance of the engine can be measured more accurately nearer to the upstream end of the turbine. The principal disadvantages of this method are that the number of thermocouples required for averaging becomes greater and the environmental temperatures in which they must operate are increased. However, as the temperature drop across the turbine varies in a known manner, it is usual to measure the temperature at the turbine outlet by locating a small number of thermocouple probes in the area of the exhaust unit, in other words by adopting position (i) In certain types of turbo propeller engine control position (ii) is adopted by locating thermocouple probes at the leading edges of intermediate nozzle guide vanes.

(b) Gas temperature indicators, referred to variously as exhaust gas temperature (EGT), turbine gas temperature (TGT) or Jet pipe temperature (JPT) indicators, may be of the semicircular or circular scale type calibrated over the ranges 0-800°C and 0-1000°C respectively. Terminal type connections are provided at the rear of the cases, the terminal identified by a positive sign being connected to the chromel lead of the cable system. Adjustment of pointers to the required datum temperature values is affected by a device which depending on the type of aircraft may be positioned either by a screwdriver or special adjusting tool.

**Thermocouple Probes**

A gas temperature thermocouple is mounted in a ceramic insulator and encased in a metal protection sheath the whole assembly forming a probe which can be projected into the gas stream. The thermocouple is made from Chromel ( a nickel-chromium alloy) and Alumel (a nickel-aluminium alloy). The hot junction protrudes into a space inside the end of the sheath which has transfer holes in it to allow the exhaust gas to flow across the hot junction. The relative positions of the transfer holes depend on whether the thermocouple is of the 'stagnation' type or the 'rapid response' type. In the 'stagnation' type which is applied to turbojet engines, the exhaust gas enters the probe through a forward facing inlet, and after circulating round the hot junction it passes through a smaller exist hole higher up and on the opposite side to the inlet hole. This arrangement allows the gas passing through to become relatively stagnant thus minimising the effects of the high velocities. The 'rapid response' type of thermocouple is designed for use in low exhaust gas velocities as in turbo propeller engines, and has equal size transfer holes arranged directly opposite each other so that gas can pass over the hot junction with minimum stagnation.

a) Probes may be of single, double or triple thermocouple element construction. A single element provides for temperature measurement only and a double element provides an additional identical circuit for transmitting a temperature signal to an engine temperature control system. The triple element type of probe provides for a further circuit for use in a warning system to detect a combustion fault. In some versions, the probe sheath is contained with a heat-resistant metal pitot tube which senses the exhaust unit pressure required when power loss indicators form part of engine instrumentation. A sleeve inside the lower end of the tube separates it from the probe sheath thus forming an annulus between the two. Three forward facing holes in the tube connect the annulus to exhaust unit pressure which is transmitted to the indicating system via a sensing line coupled to a union in the probe mounting flange, and a pressure manifold.

b) In order to obtain a good average indication of gas temperature conditions and also to ensure functioning of the indicating system in the event an element becomes defective, a number of probes are radially disposed in the gas stream and the electrical outputs are connected to form a parallel circuit. The cables from the thermocouple probes are formed as a harness around the engine and terminate at a junction box which also provides the connecting point for the cables leading to the indicator.

c) In some engine installations thermocouple probes may also be positioned in the air intake to transmit signals to a temperature indicating and control system, thus giving a reading of gas temperature compensated for intake temperature variations.

**External Circuits**

In addition to temperature/e.m.f. characteristics, the calibration of thermo-electric system indicators takes into account certain constant external circuit resistance values. The external circuit is the section from the thermocouple probe(s) to the indicator terminals and some typical resistance values are : 2 and 8 ohms for engine cylinder head temperature indicators, and 8 and 25 ohms for gas temperature indicator. In some cases the appropriate resistance values may be found on the indicators.

a) The cables connecting probes and indicators may be one of the following types:

i) Extension or those made of the same material as the thermocouple.
ii) Compensating or those which are made of a different material to the thermocouple, but having a resistance equal to that of the equivalent extension cables.

b) Whichever type lead is used a stranding is selected to make up the correct overall circuit resistance required. In some turbine gas temperature indicator installations, a trimming resistor is connected in one of the leads to adjust the external circuit resistance to the required value. The resistor is made of either Eureka (a copper-nickel alloy) or Manganin ( a copper manganese alloy) wire wound on a spool. A Eureka resistor is connected in the negative lead, while a Manganin resistor is connected in the positive lead.

c) In addition to the trimming resistor, the external circuit of an indicating system used with certain types of turbine engine includes a ballast resistor. The resistor is fitted because an engine may have an acceptable turbine inlet temperature but have a high exhaust temperature due to temperature scatter in the exhaust unit. The resistor reduces this scatter and brings the indicated temperature down. The value of the resistor is determined from test bed results, and because of this it must not be replaced by a resistor of a different value during the overhaul life of the engine. The specified resistance value is usually marked on the thermocouple system junction box or engine data plate and is also included in the engine log book.

**Engine Vibration Indicating Systems**

These systems provide continuous indication of vibration conditions normally existing when turbine engines are running, and from any sudden variations in amplitude can also provide an early warning of defect in the internal rotating parts and bearings, permitting corrective action to be taken before extensive damage occurs.

A system consists of an engine mounted pick-up unit which, in a typical application contains a spring-supported magnet and inductor coil assembly, an amplifier and an indicator mounted on the appropriate instrument panel. The scale of the indicator is graduated in units of fixed relative amplitude. A test push switch and an amber warning lamp also form part of a system. The power supply required for system operation is 115-volts single-phase, 400 Hz.

When an engine is running and electrical power is applied to the system, vibration causes relative motion between the magnet and coil and signal voltages are induced in the coil which are proportional to the velocity of vibration. These signals are applied to the amplifier where they are processed and fed to the indicator moving coil causing displacement of the pointer to positions indicating corresponding values of relative vibration amplitudes. If the vibration level exceeds a predetermined value a circuit is completed to illuminate the warning lamp. The purpose of the test switch is to check the continuity of the complete circuit. Operation of the switch connects a standard test signal, generated in the indicator-amplifier, to the pick-up unit and if the circuit is fault-free the test signal produces a standard value of vibration amplitude on the indicator.

**FUEL FLOWMETERS**

Fuel flowmeters measure and provide visual indication of the instantaneous rate at which fuel flows to an engine, and in the majority of engine fuel system installations, provide a secondary indication of the quantity of fuel consumed. A system consists of a transmitter connected in the main fuel supply line and a motor-driven indicator calibrated in either gallons, kilogrammes or pounds per hour. The transmitter contains a specially calibrated metering unit and an electrical signal pick-off unit. As fuel passes through the metering unit it drives a rotating elements which is coupled to the pick-off unit. Signals proportional to the fuel flow rate are induced in this unit and are transmitted to the indicator motor which positions the pointer at the scale graduation corresponding to the flow rate.

**Fuel Flowmeter Systems**

Fuel flowmeter systems are used to indicate fuel usage. They are most commonly installed on large multi-engine aircraft, but they may be found on any type of aircraft if fuel economy is an important factor.

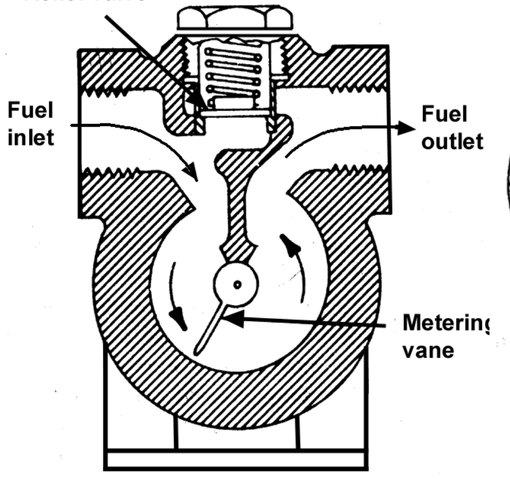



*Fig. 19.20, Flowmeter fuel chamber.*

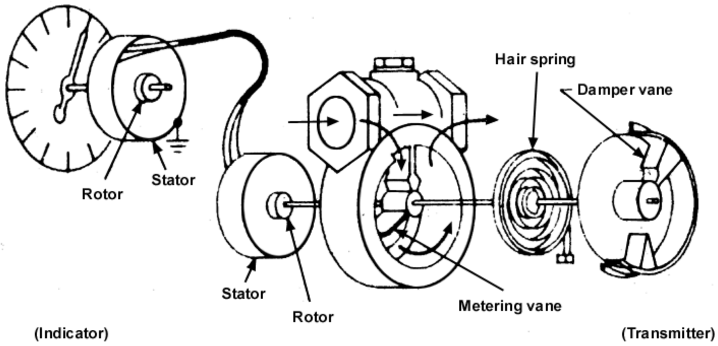



*Fig. 19.21, Fuel flowmeter system.*

A typical flowmeter system for a reciprocating engine consists of a flowmeter transmitter and an indicator. The transmitter is usually connected into the fuel line leading from the carburettor outlet to the fuel feed valve or discharge nozzle. The indicator is normally mounted in the instrument panel.

A cross sectional view of a typical transmitter fuel chamber is shown in fig. 19.20. Fuel entering the inlet side of the fuel chamber is directed against the metering vane, causing the vane to swing on its shaft within the chamber. As the vane is moved from a closed position by the pressure of the fuel flow, the clearance between the vane and the fuel chamber wall becomes increasingly larger.

Figure 19.21 shows an exploded view of a fuel flowmeter system. Note that the metering vane moves against the opposing force of a hairspring. When the force created by a given fuel flow is balanced by spring tension, the vane becomes stationary. The vane is connected magnetically to the rotor of a transmitter, which generates electrical signals to position the cockpit indicator. The distance the metering vane moves is proportional to, and measure of, the rate of fuel flow.

The damper vane of the transmitter cushions fluctuations caused by air bubbles. The relief valve bypasses fuel to the chamber outlet when the flow of fuel is greater than chamber capacity.

A simplified schematic of a vane-type flowmeter system (figure 19.22) shows the metering vane connected to the flowmeter transmitter and the rotor and stator of the indicator connected to a common power source with the transmitter.

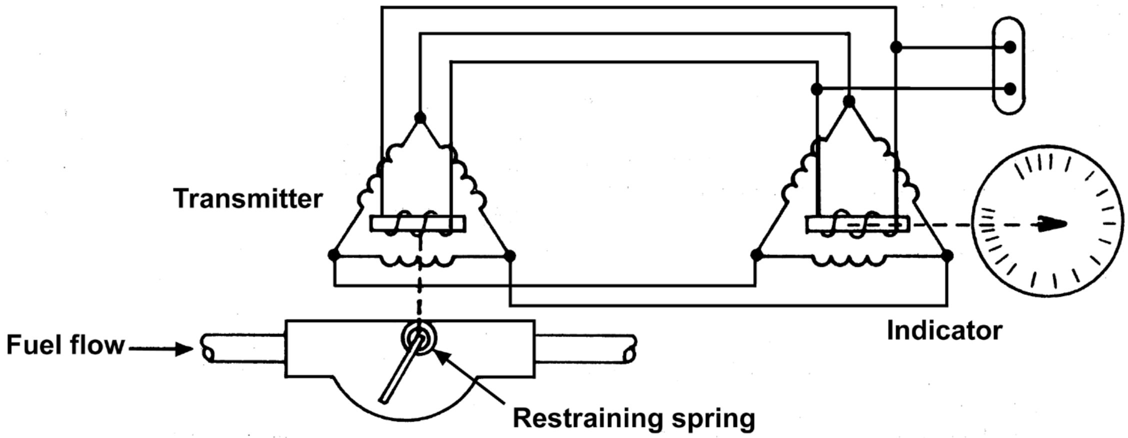



*Fig. 19.22, Schematic of vane-type flowmeter system.*

The dial of a fuel-flow indicators shown in figure 19.23. Some fuel-flow indicators are calibrated in gallons per hour, but most of them indicate the measurement of fuel flow in pounds.

The fuel flowmeter system used with turbine engine aircraft is usually a more complex system than that used in reciprocating engine aircraft.

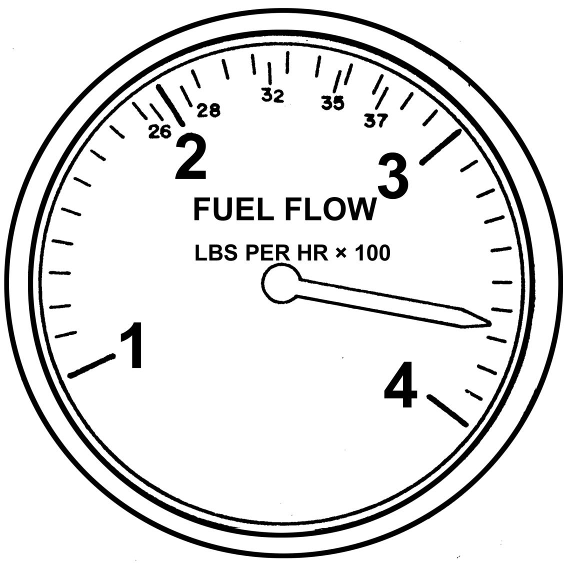



*Fig. 19.23, Typical fuel-flow indicator.*

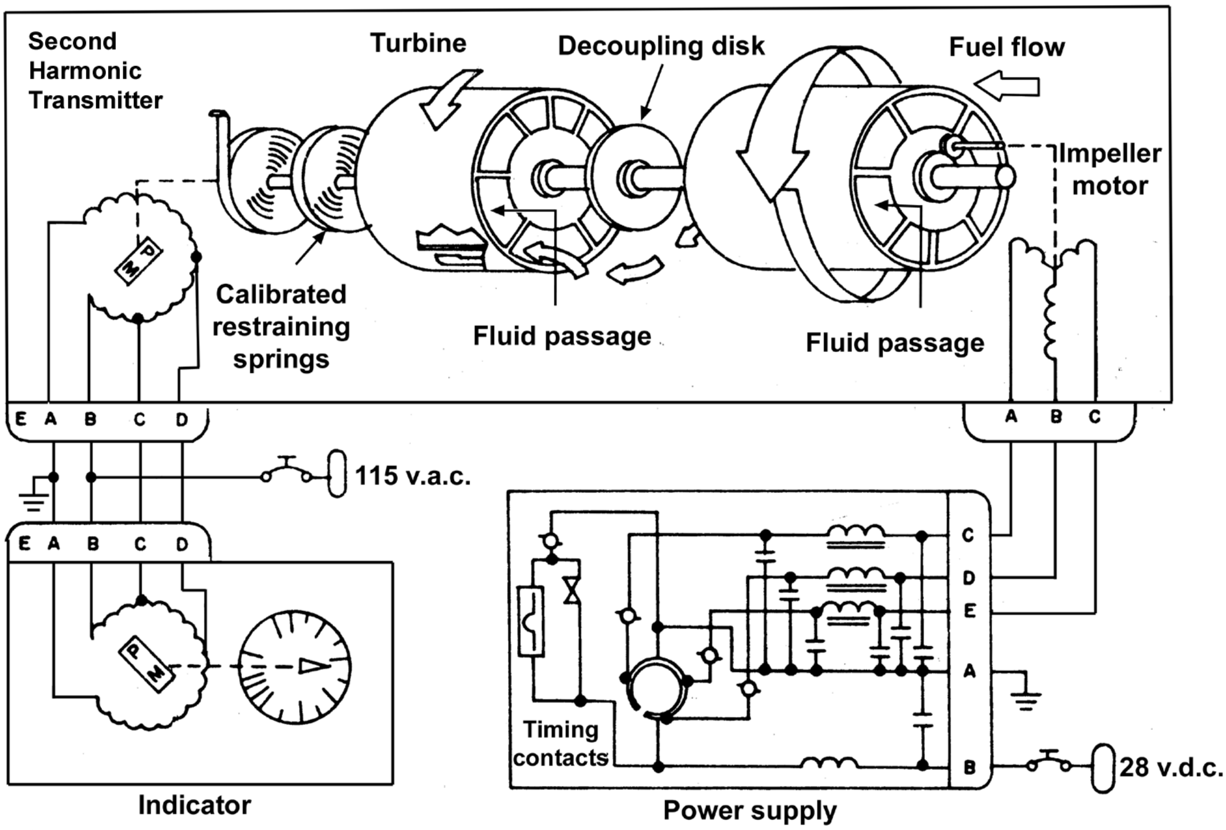



*Fig. 19.24, Schematic of a large turbine engine flowmeter system.*

In the system shown schematically in figure 19.24, two cylinders, an impeller, and a turbine are mounted in the main fuel line leading to the engine. The impeller is driven at a constant speed by a special three-phase motor. The impeller imparts an angular momentum to the fuel, causing the turbine to rotate until the calibrated restraining spring force balances the force due to the angular momentum of the fuel. The deflection of the turbine positions the permanent magnet in the second harmonic transmitter to a position corresponding to the fuel flow in the line. This turbine position is transmitted electrically to the indicator in the cockpit by means of a selsyn system.

■■■

# CHAPTER : 20
# AIRCRAFT RADIO NAVIGATION INSTRUMENTS

**MAGNETIC COMPASS**
Direct-reading magnetic compasses were the first of the many airborne flight and navigational aids ever to be introduced in aircraft. Their primary function is to show the direction in which an aircraft is heading with respect to earth's magnetic meridian.

As far as present-day aircraft and navigational aids are concerned, however, a subdivision of this function has been brought about by the type of aircraft and by the aids employed. For example, in many small aircraft the magnetic compass is utilized as the primary heading indicator, while in aircraft employing remote-indicating compasses and other advanced navigational aids, it plays the role of stand-by heading indicator.

The principle on which they operate is the very basic one of reaction between the magnetic field of a suitably suspended permanent magnet and the field surrounding the earth.

**INSTRUMENT LANDING SYSTEM (ILS)**
This is a system which aids a pilot in maintaining the correct position of his aircraft during the approach to land on an airport runway. Two radio signal beams are transmitted from the ground; one beam is in the vertical plane and at an angle to the runway to establish the correct approach or glide slope angle, while the other, known as the localizer, is in the horizontal plane; both are lined up with the runway centre-line.

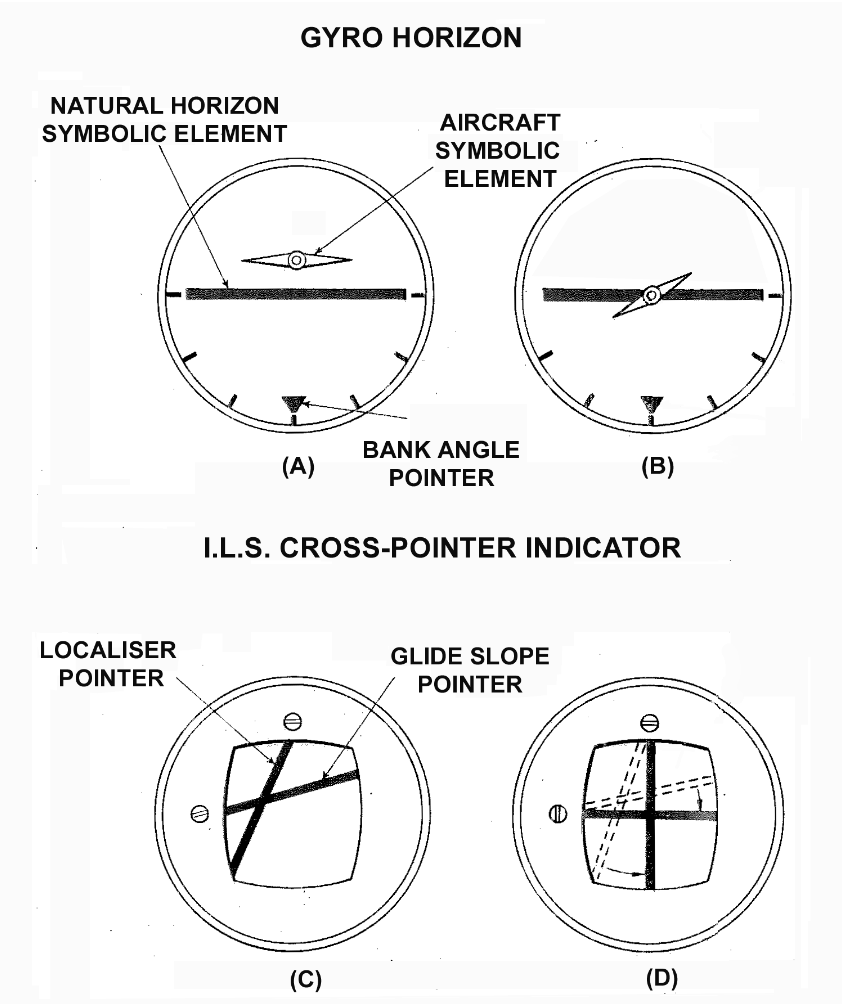



*Fig. 20.1, Examples of director display (A) 'Fly down' directive; (B) 'bank right' directive; (C) 'Fly left' and 'Fly up' directive; (D) response matches directive.*

A receiver on board the aircraft receives the signals and transmits them to a cross-pointer type of instrument on the main instrument panel. Meters within the instrument are monitored by the glide slope signals and localizer signals, and respectively control horizontal and vertical pointers.

When the aircraft is on the approach to land and is, say, below the glide slope beam, the glide slope (horizontal) pointer of the instrument will be deflected upwards as shown in Fig. 20.1 (C). Thus, the pilot is directed to "fly the aircraft up" in order to intercept the beam. Similarly, if the aircraft is to the right of the localizer beam the vertical pointer will be deflected to the left thus directing the pilot to "fly the aircraft left". As the pilot responds to the instruments directives the pointers move back to their centre positions, indicating that the aircraft is in the correct approach position for landing (Fig. 20.1(D)

It will be apparent from the diagram that as the aircraft is manoeuvred in response to demands, the pointer movements; for example, in responding to the demand "fly left" the vertical pointer will move to the right. However, in turning to the left the bank attitude of the aircraft will change into the direction of the turn, that is, the left wing will go down, and as the gyro horizon will indicate this directly, then by monitoring this instrument the pilot can crosscheck that he is putting the aircraft into the correct attitude in responding to the demands of the ILS indicator.

**AUTOMATIC DIRECTION FINDER (ADF)**

(ADF) Automatic Direction Finder is a very useful aid to aerial navigation. It can be used for homing to or finding direction from any station that broadcasts in low and medium frequency radio bands.

Aviation installation that operate in low frequency band include non-directional beacons. Radio range station, automatic direction finder is also able to receive commercial AM broadcasting station and home on them.

**Automatic Direction Finder Equipment**

It consists of an automatic direction finder receiver, a loop antenna, a sense antenna and a bearing indicator.

**Antenna**

Two antenna are required for ADF operation. One of these known as sense antenna is a non directional antenna that has the capability of providing directional information.

The other antenna is a loop antenna which senses magnetic bearing from airplane to station.

The sense antenna usually is a long wire installed on top of airplane stretching from an insulator near the top of fuselage back to stabilizer.

The loop antenna, a metal ring enclosing coils of insulated wires is usually contained within a streamlined brushing mounted well forward on underside of fuselage.

**The ADF Receiver**

Modern ADF receivers are digitally tuned providing rapid and precise tuning to any desired frequency. They receive stations throughout low and medium frequency band.

Modern ADF receivers are digital tuned provided rapid and precise tunning and desired frequency they receive station throughout low and medium frequency band ADF receiver like Bendix ADF receives information from antenna and translates it into movement of needle of bearing indicator.

**Bearing Indicator**

An ADF bearing indicator incorporates a bearing pointer and an azimuthal dial which is graduated in degrees from 0 to 360$^{o}$. Bearing pointer always points in direction of station to which receiver is tuned.

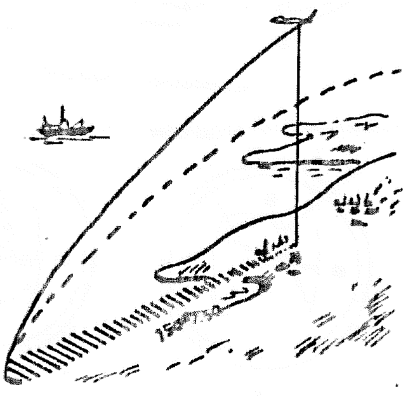


*Fig. 20.2, Illumination of the Ground Surface by Aircraft Radar.*

Some indictors have a fixed card, others have moveable cards with a bearing indicator with a fixed card, most of airplane is always oriental to $0^{0}$ or bearing indicator regardless of actual magnetic heading of airplane.

With a bearing indicator, with movable card, pilot is able to rotate the azimuthal card so that it corresponds with DG heading of the airplane.

ADF bearing indicator (fixed card) indicates that station to which ADF receiver is lined.

An ADF bearing indicator with a movable azimuthal card automatically displays magnetic bearing of station.

**DISTANCE MEASURING EQUIPMENT (DME)**
The purpose of DME is to give pilot a continuous reading of his distance, to a fixed ground station. This ground installation is usually a VORTAC or a TACA station transmitting a radar beacon on VHF.

Airborne equipment transmits pulse to ground station, that in effect ask " How far I am from your station". Each aircraft interrogation has a unique transmission rate or pattern which found station reproduces in its replies to an individual aircraft from station. Airborne DME receiver selects the supply pulses designated for itself (i.e. those with same rate relationship as used in its own original transmission), measures electronically time intervals between its own transmission and received reply, converts elapsed time into a mileage fig which is continuously displayed in cockpit indicator as slant range from aeroplane to the station. The range increases as plane flies away from a station and decreases as it approaches the station.

Slant range can be converted into found distance by following formula. Ground distance is square of $S^2$ (Slant range in nautical miles) minus $A^2$ (altitude in nautical miles), when slant range is more than double the altitude the difference between slant range and ground distance will be appreciable. In such a case, pilot must employ above formula to determine his actual ground distance from station. Unless his DME includes a component solver that converts slant mileage to horizontal mileage automatically.

Most DME units also have capability to give a ground speed reading. All DME distance readings are in nautical miles and ground speed readings are in knots.

Ground speed features of DME is of essential value to pilot for it allows him to choose altitude where he will encounter most favourable winds. By flying at several different altitudes and observing ground speed read out on DME, he can determine at which altitude ground speed was highest thereby indicating most favourable winds. The DME is also used in making orbiting or arc approaches to the airport.

DME operates in 960 to 1215 MHz UHF range, transmitting in range 960 to 1024 MHz and 1151 to 1213 MHz and receiving in 1025 to 1150 MHz range.

DME equipment is considered reliable upto 100 miles from station and so accurate within about 2%.

**Radars**
Radio location, using the property of radio waves of being reflected by the ground and objects (bridges, roads, cities, etc. makes it possible to obtain a radar image and determine the distance (range) to these objects (Fig. 20.2)

The phenomenon of reflection of radio waves (from sea vessels) was observed for the first time by the inventor of radio in 1897.

The distance is measured as follows: A transmitting device radiates radio waves in short quantums called main pulses. Every main pulse with a velocity equal to that of light reaches the object whose distance is to be determined and is reflected from it.

A large number of experiments carried out by scientists showed that the velocity of propagation of radio waves in space depends very little on the ground conditions, of the atmosphere, season or time of day and is a constant quantity, roughly equal to 300,000 km/sec. The reflected radio pulse, or as it is called, the "blip" returns with the same speed to the transmitter. If a receiver is installed near it then with its help it is possible to receive the blip.

In order to determine the distance it is sufficient to know the time lapse between the moment of transmitting the main pulse and that of receiving the blip.

In radar the time is measured by an electric method in a special device. The distance is measured directly from the screen of a cathode ray tube similar to that of a common television set in its working principle and external appearance.

On this indicator signs in the form of light sports are illuminated. They correspond to the main pulse and blip. Using the scale plotted on the screen to the time scale the distance in kilometres to the object is determined by the distance between these signs.

The signs from different objects situated in the same direction but at different distances from the radar are located on the screen in one straight line but at different distances.

It is often sufficient to determine the distance to the object if the object is identified. But sometimes it is necessary to obtain the image of the ground or sea surface along with the objects situated on it, or in other words a radar map of the place is necessary.

Obtaining a radar image is based on an important property of radio waves, which are not reflected identically from different objects. The magnitude of the reflected signal varies in accordance with the size and properties of the reflecting surface.

Radio waves are well reflected by land, ships, iron bridges and structures and are least reflected from a water surface. The better the radio reflecting ability of the object the larger is the magnitude of the blip. A large blip illuminates a brighter spot on the indicator screen. Therefore objects reflecting the radio waves will create on the screen a spot brighter than the blip from a water surface.

Let us assume that it is necessary to search for ship at sea. We will move the direction of the radiating main pulses in the horizontal plan (along with the azimuth). These pluses will then successively illuminate first the sea, then the ship and finally the sea again. The blips from the sea surface will show up as a weak light spot on the screen in comparison with the blips from the ship. Therefore on the screen there will appear a bright spot on a light background-the ship sought.

It is necessary to note that radar for detection of objects radiates not one but a number of main pulses in succession and obtains the same number of reflected ones. The fact is that the energy of the blips is very small and in order to obtain a clear image on the indicator screen it is necessary to deliver on it up to ten and sometimes many more signals for one object sought. The next main pulse is not radiated by the transmitter unit until the receiver receives the blip from the previous main pulse. Otherwise superimposition of signals takes place and the radar receiver cannot work.

**Antenna**

A radar antenna consists of source of radio radiation (radiator) and a reflecting mirror (reflector). The radio beams from the radiator, after falling on the reflector, are reflected from it according to laws of optics and "illuminates" as it were, a narrow sector of the locality.

The geometrical dimensions and shape of the reflector for a given radiated wavelength determine the angular dimensions and shape of the sector where radar radiation is being directed. The zone can be obtained graphically.

Usually two graphs are given, one in a horizontal plane and the other in the vertical one. These graphs are called radiation pattern in the horizontal and vertical planes respectively.

Depending on the purpose of radars their antennas have various radiation patterns. For panoramic radar used on aircraft to obtain a radar map of the site the radiation pattern resembles a wide fan directed toward the ground. The width of this "fan" in the horizontal plane varies from fractions of degree to several degrees. In the vertical plane it is wide and the width is in tens of degree.

Designers always strive to obtain a narrow beam because the narrower the beam the finer the objects on the site that a radar can "see" or, as the specialists say, a radar with a narrow beam has a good resolving power.

The antenna rotates at speed of some tens of rotations per minute with the help of an electric drive. During its rotation the radio beam "sweeps" a considerable portion, can reach 150-250 km.

The impulse principle of operation of radar makes it possible to use one and the same antenna to transmit main pulses and receive the blips. This is very important since to install two antennas of 1-1.5m size each on an aircraft would be an extremely difficult task.

The antenna hurts the aerodynamic properties of the aircraft and occupies much useful space. It must never be accommodated inside the aircraft since the metallic fuselage would absorb the radio waves. It therefore becomes necessary to accommodate the antenna on the exterior of the aircraft in a special housing (cowl) of material transparent to radio waves.

A cowl with the antenna is mounted in the aircraft's nose or under its fuselage. To reduce the drag it is made in the form of hemisphere or half a pear.

**Transmitting Equipment**

It is designed to generate high-frequency main pulses with the frequency corresponding to the operating wave of radar (aircraft radar usually operates in the centimetre wave band).

The power of high-frequency pulses or, as it is called, the power of the transmitter, must be very large. This is due to the fact that electromagnetic radiation of radio pulses is absorbed and partially dissipated by thermosphere. Besides this is very important for the reliable operation of radar to obtain blips of sufficient strength, for which it is necessary to raise the strength of the main pulses.

The object to be detected by radar are illuminated by pulses whose strength is measured by tens of kilowatts. But the strength of the blips received is very often considerably less than one billionth of a watt.

The transmitter generates a high frequency pulse of large strength only for a very short interval of time equal to 0.5 - 1.5 micro-secs. Depending on the range of radar the transmitter "gives out' some hundreds to some thousands of such pulses in 1 sec. In the intervals between the pulses the transmitters as it were "takes a rest". Storing up energy from a relatively low-powered supply source.

The high-frequency pulses in transmitting equipment are produced by special electro vacuum valves known as magnetrons. This takes place as follows: inside a copper vacuum chamber of the magnetron a cylindrical cathode is placed. The chamber, whose inner part usually has the shape of a ring with figured notches, itself acts as the anode. The magnetron is situated between the poles of a powerful magnet whose field also takes part in creating high-frequency oscillation inside the magnetron.

From a special device called a modulator high voltage pulses enter the magnetron. Within 0.5-1.5 micro-sec of the action of this voltage the magnetron produces (generates) high frequency main pulses which then go to the antenna radiator through a wave guide. The transmitting equipment is a separate block pressurised by air pump. Pressurisation is essential because at high altitudes where the atmospheric pressure is small the insulation properties of the air are bad. Under such conditions high voltage electrical breakdown and failure of transmitting equipment are possible.

**Receiving Equipment**

The electromagnetic energy reflected from objects is received by an antenna and comes to the radar receiver through a wave guide, However, the signals received cannot be delivered directly to the indicator screen as they are very weak. It is necessary to amplify them and convey them into video pulses preserving only the shape of the signals received. Such pulses are free of the high-frequency oscillation which are received by the antenna of the transmitter.

The device that convert a high-frequency signal into video signal is called detector.

Since the strength of the blip is small the sensitivity (receptiveness) of the receiver must be quite high. For the sake of comparison we will note here that modern television sets have a sensitivity (receptiveness) some hundred times inferior.

Let us briefly study the peculiarities of radar receiving equipment. In such devices, as a rule, a circuit making possible a large amplification of the signals received is used. For this purpose the high-frequency signals received is mixed with the signal from an additional generator (oscillator). The mixing takes place both in the mixer diode and as in the usual mixer of a broadcasting radio.

At the exit of the mixer a pulse voltage of intermediate frequency usually 30 to 60 MHz, is delivered. A voltage of such frequency can be amplified comparatively easily. For the purpose of amplification there are six to eight cascades of intermediate frequency amplifiers.

The amplified pulses enter the detector (crystal diode) where from the high-frequency pulse its envelop is given out as a video signal.

Super oscillating radar receivers differ from broadcasting and television receivers by the fact that as an oscillator a special electro-vacuum, device known as klystron is used, while a crystal diode is used as a mixer.

From outside a klystron resembles a common radio valve. It is a generator of auxiliary high frequency oscillations used for conversion (with the help of a mixer) of blips into pulse oscillations of intermediate frequency.

The klystron is used in radar not only due to its ability to generate high frequency oscillations Even an ordinary radio valve can produce such oscillations. The most important property of the klystron is to vary the frequency of the oscillations generated on varying the voltage at its anode (reflector electrode). This property of the klystron makes it possible to use an automatic frequency control system in the receiver and thus change the tunning of the receiver without the operator's intervention.

The necessity of receiver control arises due to insufficient operations stability of the radar transmitting equipment. The fact that the magnetron of the transmitter, under the influence of a number of factors, arbitrarily varies its frequency and, consequently, the frequency of the blips.

If the receiver has fixed tuning while the frequency of the signals received varies, then no stability of reception of blips is possible. Therefore an automatic frequency control system is used in the receiver, making possible the automatic control of the frequency of the klystron in accordance with the variation of blip frequency.

**Synchronizing And Indicating Devices**

In order to determine the distance objects a "clock" that can measure time with an accuracy up to some microseconds is needed in the radar. Such accurate measurements have become possible due to the uses of crystal clocks.

A crystal plate placed in a variable electrical field is capable of creating in the oscillating circuit of high-frequency generator oscillations of particular frequency directly proportional to the geometric dimensions of the plate.

This property of crystal made it possible to built generator with the help of which a very accurate measurement of time was achieved. The high frequency oscillations of such a generator are converted into a sequence of electrical pulses, one following the other at equal intervals of time for instance 13.333 microseconds. By this time the main pulse would travel 2 km and, having been reflected, return (1 micro-sec in radar corresponds to 150 m).

By delivering "commands" to the transmitter, modulator, receiver and indicator the crystal clock makes them operate synchronously and creates on the indicator screen scale marks by which the distance of the objects detected is determined.

The basic indicating device is cathode-ray tube. The deflector of this tube makes the electron ray traverse its screen in the required direction.

While traversing it under the influence of the deflection yoke the ray forms something like a grid-sweep array. Sometimes the deflection yoke causes the ray to move on the screen along other paths, for instance from the centre along \the radius. The ray begins to move along the radius of the screen (radial sweep) with a signal from the crystal clock simultaneously with the radiation of the main pulse and, at the rime of radiation of the next main pulse, completes its movement toward the tube rim (approximately after 500 to 2,000 microseconds depending on the set scale of radar range).

Simultaneously with the radial movement of the ray the deflection yoke turns the sweep about the tube axis in synchronization with the rotating antenna (nearly 20 rotations/min). Due to this dual motion of the ray a circular array occupying the whole area of the screen is obtained.

The blips enter a controlling electrode of the tube from the receiver and, depending on their magnitude, amplify the electronic beam, proportionally increasing the illumination of the screen and spotlighting the objects from which they were reflected.

The electric signals of time from the crystal clock are also constantly fed to the controlling electrode. These signals create on the screen bright concentric scale rings located at an equal distance from each other. The distances of the detected objects are determined from these rings.

**The Applications of Radar**

Besides the basic task of survey of the ground surface many other navigation tasks are also laid on aircraft radar.

Radar enables the crew to detect mountains and thunderclouds. By observing thunder fronts the pilot can choose zones (corridors) where thunder storm activity is weakest and does not obstruct the safe flight of an aircraft. Besides this radar can be used for operation with ground and aircraft transponder-beacons. On the indicators of the aircraft radar the ground transponder-beacons, whose locations is known in advances, are illuminated by bright spot. The spots from the transponders flicker. According to the frequency of flickering the location of the transponder is determined with respect to the location of the aircraft.

In cases where the danger of aircraft collision arises radar warns the pilot by switching on a special danger signals light.

The transponder-beacon consists of a receiver and a pulse transmitter. The receiver and the transmitter are tuned to the operational frequency of the aircraft radar. The receiver outlet is connected to device that controls the operation of the transmitter.

If an aircraft radar sends a main pulse and the receiver of a transponder beacon receives it, a controlling signal from the receiver outlet is fed to the transmitter and makes it radiate an answer signal (something like a blip).

The aircraft receiver receives the signal and delivers it to the radar indicator screen where it places a mark whose range and azimuth correspond to the distance and bearing between the transponder-beacon, the signal of the transponder-beacon transmitter is much more powerful. This considerably increases the radar range in working with transponder-beacons and eases the navigator's work.

Transponder-beacons built on the same principles are installed on aircraft to increase their reliability and range of detection of their aircraft equipped with radar.

■■■

# CHAPTER : 21
# AUTO PILOT

**AUTOPILOT SYSTEM**

The automatic pilot is a system of automatic controls which holds the aircraft on any selected magnetic heading and returns the aircraft to that heading when it is displaced from it. The automatic pilot also keeps the aircraft stabilized around its horizontal and lateral axes.

The purpose of an automatic pilot system is primarily to reduce the work, strain, and fatigue of controlling the aircraft during long flights. To do this the automatic pilot system performs several functions. It allows the pilot to manoeuvre the aircraft with a minimum of manual operations. While under automatic control the aircraft can be made to climb, turn, and dive with small movements of the knobs on the autopilot controller.

Autopilot systems provide for one, two, or three axis control of the aircraft. Some autopilot systems control only the ailerons (one axis), others control ailerons and elevators or rudder (two axis). The three-axis system controls ailerons, elevators, and rudder.

All autopilot systems contain the same basic components : (1) Gyros, to sense what the airplane is doing ; (2) servos, to move the control surfaces ; and (3) an amplifier, to increase the strength of the gyro signals enough to operate the servos. A controller is also provided to allow manual control of the aircraft through the autopilot system.

**Principle of Operation**

The automatic pilot system flies the aircraft by using electrical signals developed in gyro-sensing units. These units are connected to flight instruments which indicate direction, rate-of-turn, bank, or pitch. If the flight attitude or magnetic heading is changed, electrical signals are developed in the gyros. These signals are used to control the operation of servo units which convert electrical energy into mechanical motion.

The servo is connected to the control surface and converts the electrical signals into mechanical force which moves the control surface in response to corrective signals or pilot commands. A basic autopilot system is shown in figure 21.1.

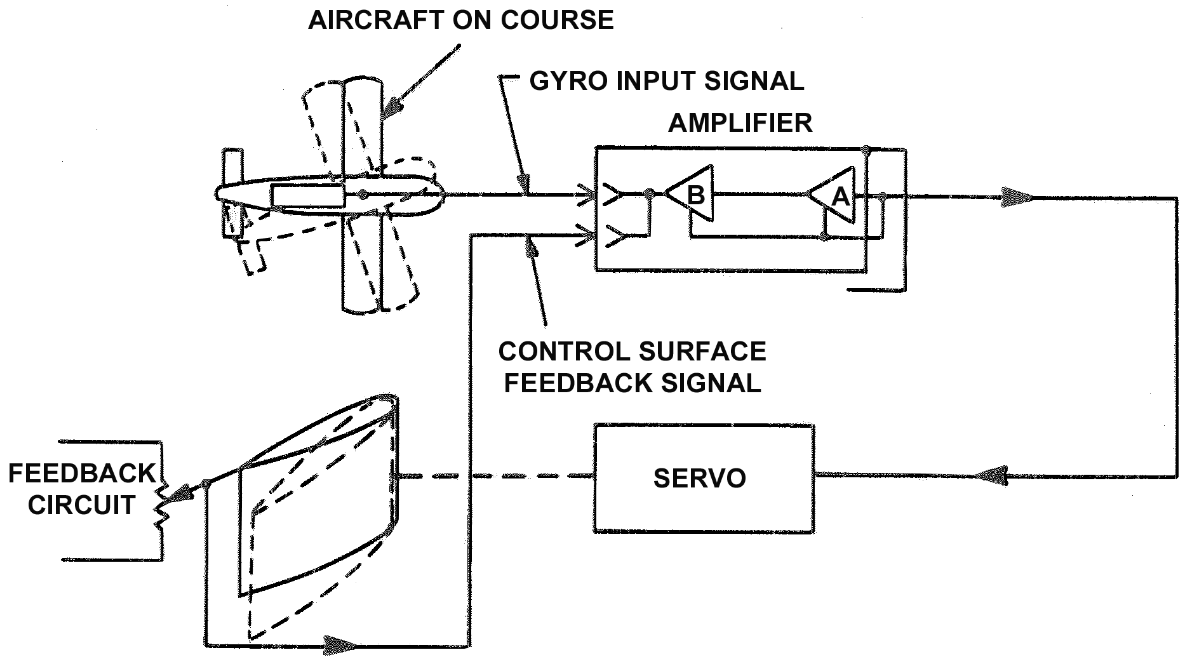



*Fig. 21.1, Basic autopilot system.*

Most modern autopilots can be described in terms of their three major channels : (1) The rudder, (2) aileron, and (3) the elevator channels.

The rudder channel receives two signals that determine when and how much the rudder will move. The first signal is a course signals derived from a compass system. As long as the aircraft remains on the magnetic heading it was on when the autopilot was engaged, no signal will develop. But any deviation causes the compass system to send a signal to the rudder channel that is proportional to the angular displacement of the aircraft from the preset heading.

The second signal received by the rudder channel is the rate signal, which provides information any-time the aircraft is turning about the vertical axis. This information is provided by the turn-and-bank indicator gyro. When the aircraft attempt to turn off course, the rate gyro develops a signal proportional to the rate of turn, and the course gyro develops a signal proportional to the amount of displacement. The two signals are sent to the rudder channel of the amplifier, where they are combined and their strength is increased. The amplified signal is then sent to the rudder servo. The servo will turn the rudder in the proper direction to return the aircraft to the selected magnetic heading.

As the rudder surface moves, a followup signal is developed which opposes the input signal. When the two signals are equal in magnitude, the servo stops moving. As the aircraft arrives on course, the course signal will reach a zero value, and the rudder will be returned to the streamline position by the followup signal.

The aileron channel receives its input signal from a transmitter located in the gyro horizon indicator. Any movement of the aircraft about its longitudinal axis will cause the gyro-sensing unit to develop a signal to correct for the movement. This signal is amplified, phase-detected, and sent to the aileron servo which moves the aileron control surfaces to corrected for the error.

As the aileron surfaces move, a followup signal builds up in opposition to the input signal. When the two signals are equal in magnitude, the servo stops moving. Since the ailerons are displaced from streamline, the aircraft will now start moving back toward level flight with the input signal becoming smaller and the follow-up signal driving the control surfaces back towards the streamline position. When the aircraft has returned to level flight in roll attitude, the input signal will again be zero. At the same time the control surfaces will be streamlined, and the follow-up signal will be zero.

The elevator channel circuits are similar to those of the aileron channel, with the exception that the elevator channel detects changes in pitch attitude of the aircraft. The circuitry of all three channels can be followed by referring to the block diagram in figure 21.2

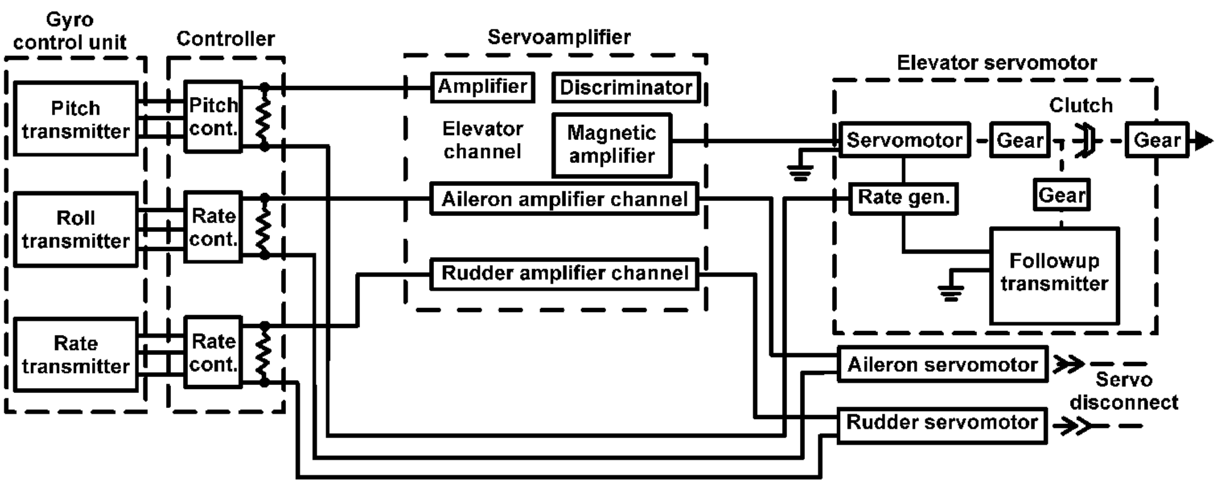



*Fig. 21.2, Autopilot block diagram.*

The foregoing autopilot system description was used to show the function of a simple autopilot. Most autopilots are far more sophisticated; however, many of the operating fundamentals are similar. Autopilot systems are capable of handling a variety of navigational inputs for automatic flight control.

**BASIC AUTOPILOT COMPONENTS**

The components of a typical autopilot system are illustrated in figure 21.3. Most systems consist of four basic types of units, plus various switches and auxiliary units. The four types of basic units are :

(1) The sensing elements, (2) command elements,

(3) output elements, and (4) the computing element.

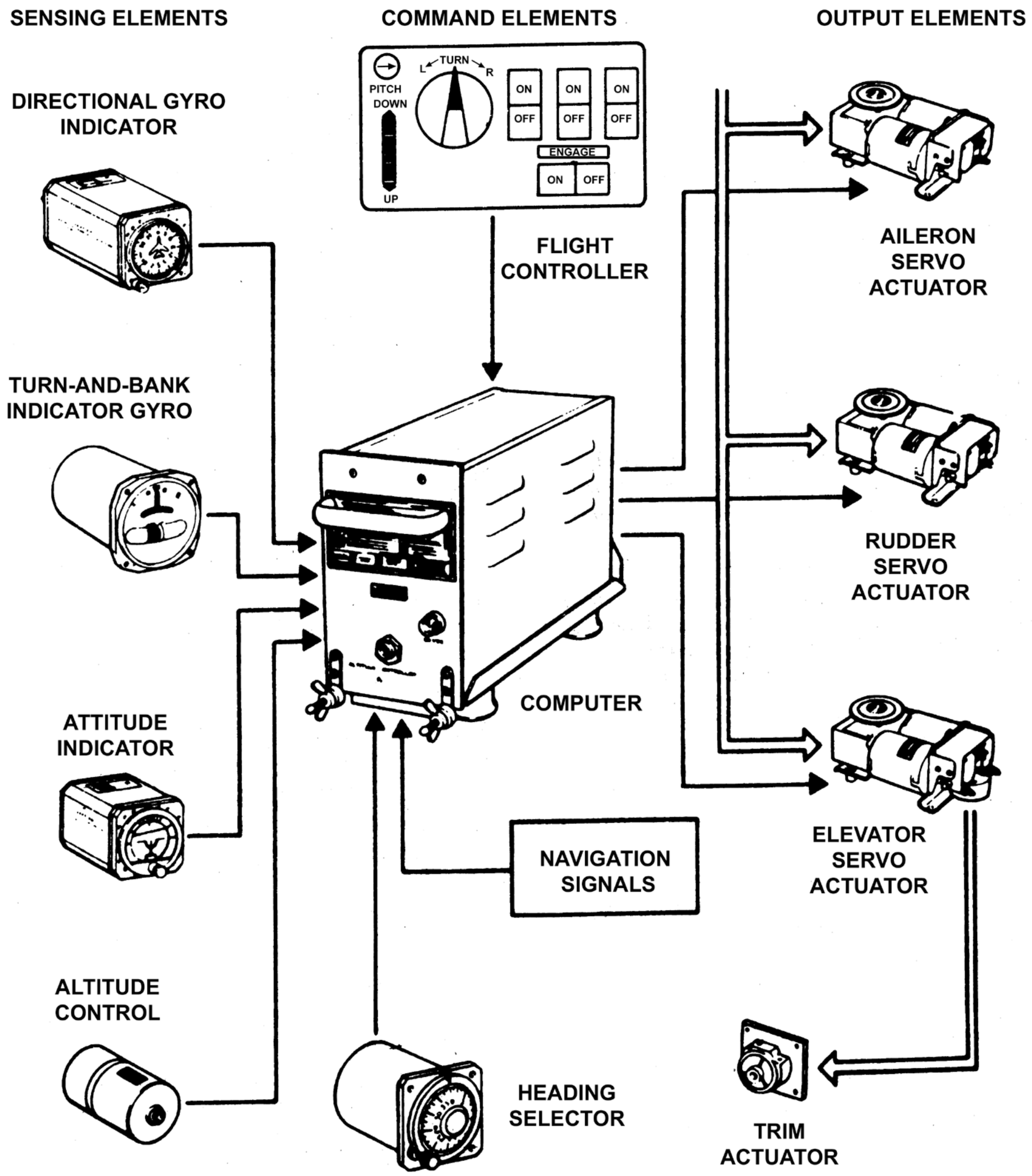



*Fig. 21.3, Typical autopilot system components.*

**Command Elements**

The command unit (flight controller) is manually operated to generate signals which cause the aircraft to climb, dive, or perform coordinated turns. Additional command signals can be sent to the autopilot system by the aircraft's navigational equipment. The automatic pilot is engaged or disengaged electrically or mechanically, depending on system design.

While the automatic pilot system is engaged, the manual operation of the various knobs on the controller (figure 21.4) manoeuvre the aircraft. By operating the pitch trim wheel, the aircraft can be made to climb or dive. By operating the turn knob, the aircraft can be banked in either direction. The engage switch is used to engage and disengage the autopilot. In addition, most systems have a disconnect switch located on the control wheel (s). This switch, operated by thumb pressure, can be used to disengage the autopilot system should a malfunction occur in the system.

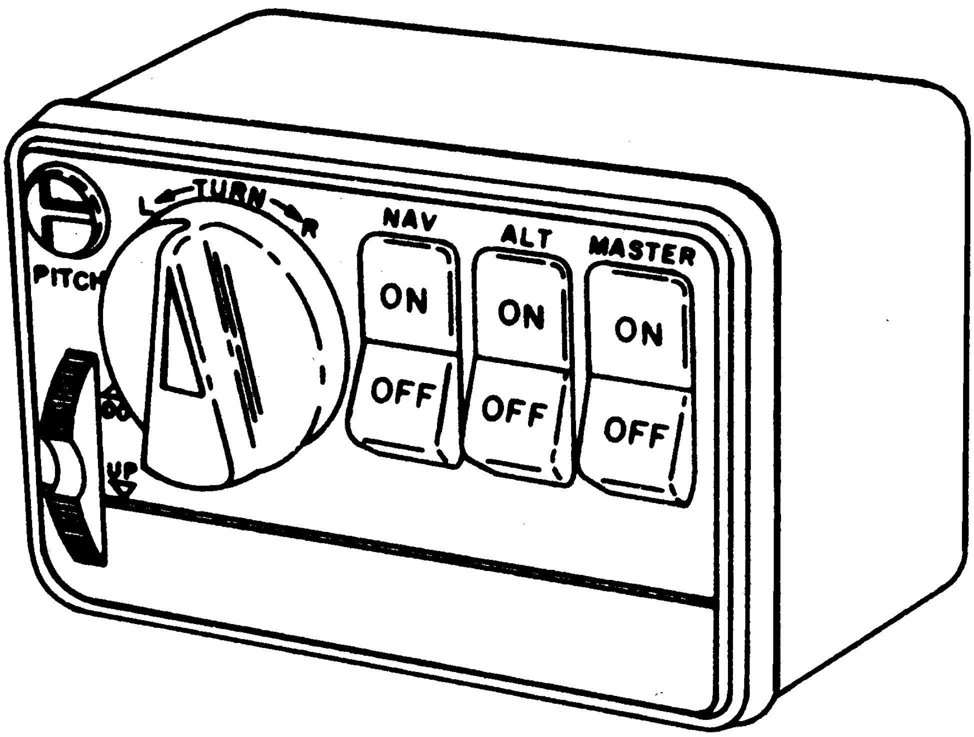



*Fig. 21.4, Typical autopilot controller.*

One type of automatic pilot system has an engaging control that manually engages the clutch mechanism of the servomotor to the cable drum. A means of electrically disengaging the clutch is provided through a disconnect switch located on the control wheel (s).

**Sensing Elements**

The directional gyro, turn-and-bank gyro, attitude gyro, and altitude control are the sensing elements. These units sense the movements of the aircraft, and automatically generate signals to keep these movements under control.

**Computer or Amplifier**

The computing element consists of an amplifier or computer. The amplifier receives signals, determines what action the signals are calling for, and amplifies the signals received from the sensing elements. It passes these signals to the rudder, aileron, or elevator servos to drive the control surfaces to the position called for.

**Output Elements**

The output elements of an autopilot system are the servos which actuate the control surfaces. The majority of the servos in use today are either electric motors or electro/pneumatic servos.

An aircraft may have from one to three servos to operate the primary flight controls. One servo operates the ailerons, a second operates the rudder, and a third operates the elevators. Each servo drives its associated control surface to follow the directions of the particular automatic pilot channel to which the servo is connected.

Two types of electric motor-operated servos are in general use. In one, a motor is connected to the servo output shaft through reduction gears. The motor starts, stops, and reverses direction in response to the commands of the gyros or controller. The other type of electric servo uses a constantly running motor geared to the output shaft through two magnetic clutches. The clutches are arranged so that energizing one clutch transmits motor torque to turn the output shaft in one direction ; energizing the other clutch turns the shaft in the opposite direction.

The electro/pneumatic servos are controlled by electrical signals from the autopilot amplifier and actuated by an appropriate air pressure source. The source may be a vacuum system pump or turbine engine bleed air. Each servo consists of an electro/magnetic valve assembly and an output linkage assembly.

■■■

# CHAPTER : 22
# ELECTRONIC (CRT) DISPLAYS

Displays of this type, which are based on the electron beam scanning technique, have been in use in aircraft for many years. For example, during World War II military aircraft used equipment developed from the then existing ground-based radar systems. With the aid of such equipment, and depending on an aircraft's specific operational role, crews were able to navigate by 'radar mapping' of terrain, to identify ground target areas, and also to detect the positions of hostile intercepting aircraft.

As far as civil aircraft are concerned, this display technology first came into prominence in 1946, with the introduction of weather radar systems to satisfy the operational requirements for transport category aircraft, and it has continued to be an essential part of the 'avionics fit' of this and other categories of aircraft.

The situation, however, of a weather radar display indicator remaining as an isolated item of video equipment was to undergo considerable change, largely as a result of systems analysis, exploration of the versatility of the CRT, and also investigation into methods whereby not only weather data, but also that associated with the many other utilities systems of an aircraft, could be programmed into computers. These had reached such high levels of sophistication and capacity for data processing that it became possible for a single CRT display unit, under microprocessor control, to project the same quantity of system status data which would otherwise have to be displayed by a very large number of conventional-type instruments. Furthermore, the introduction of CRTs and circuits capable of producing a wide range of colours made it possible to differentiate between significant parts of a display, and in particular, to lay emphasis on information of an advisory, cautionary, or warning nature.

The development of such multi-data display technology for both civil and military aircraft was also influenced by the fact that by integrating all computers via a data 'highway' bus, the scene was set for the management of all aspects of in-flight operation to be fully automated while still enhancing flight safety. This also led to improvements in levels of systems' redundancy, changes in the layouts of transport aircraft flight decks, and a reduction in crew complement with the attendant changes in their role and workloads.

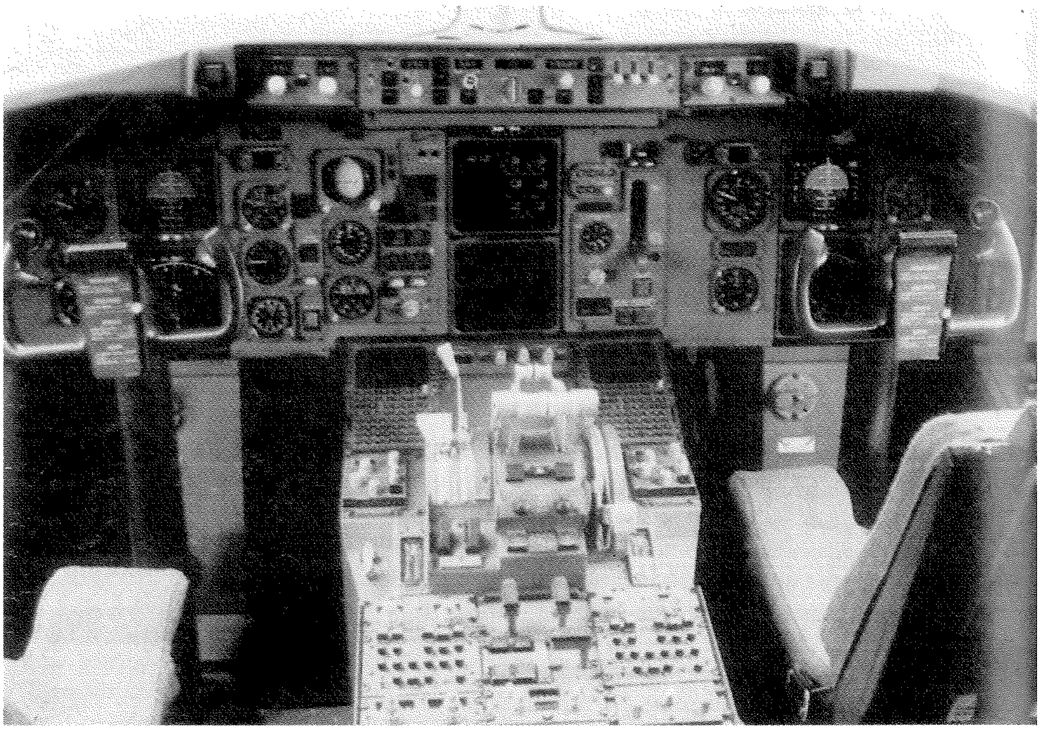


*Fig.22.1. Flight deck layout of the Boeing 757.*

The first of the 'new technology' transport aircraft (generally dubbed as 'glass cockpit' aircraft) were the Boeing 757,767 and Airbus A310. All three were launched as design projects in 1978, and both the B757 and B767 first entered commercial service in the US in December 1982. The first A310 services were operated by two European airlines in April 1983. These aircraft, and several of their descendant types, are now in service world-wide, together with many types of smaller aircraft, including helicopters, in which the foregoing technology has also satisfied an operational need.

Figures 22.1 and 22.2 show the flight deck layouts and CRT display locations of the B757 and A310 respectively.

**PRINCIPLE OF THE CRT**

A CRT is a thermionic device, i.e. one in which electrons are liberated as a result of heat energy. As may be seen from Fig. 22.3, it consists of an evacuated glass envelope, inside which are positioned an electron 'gun' and beam-focusing and beam-deflection systems. The inside surface of the screen is coated with a crystalline solid material known as a phosphor. The electron 'gun' consists of an indirectly-heated cathode biased negatively with respect to the screen,

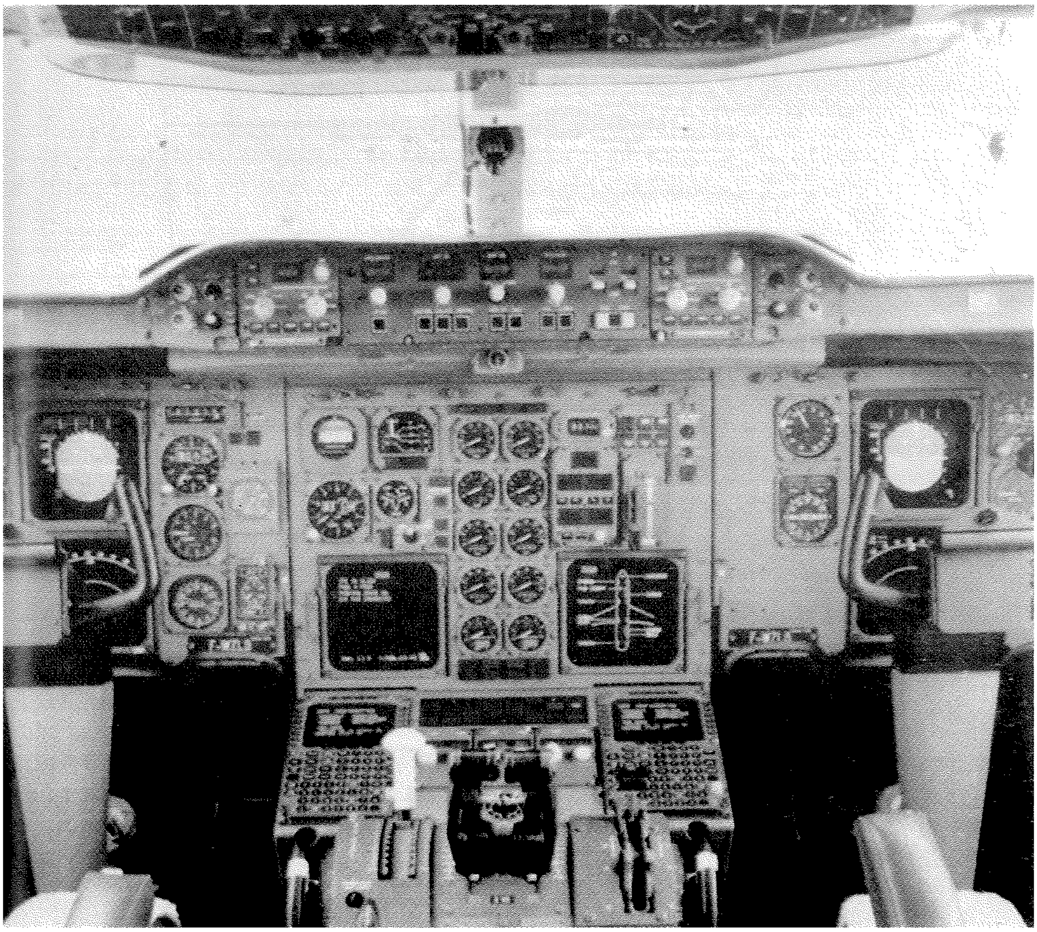


*Fig.22.2. Flight deck layout of the Airbus A 310.*

a cylindrical grid surrounding the cathode, and two (sometimes three) anodes. When the cathode is heated, electrons are liberated and in passing through the anodes they are made to form a beam.

The grid is maintained at a negative potential, its purpose being to control the current and so modulate the beam of electrons passing through the hole in the grid. The anodes are at a positive potential with respect to the cathode, and they accelerate the electrons to a high velocity until they strike the screen coating. The anodes also provide a means of focusing, which, as will be noted from Fig. 22.3, happens in two stages.

The forces exerted by the field set up between the grid and the first anode bring the electrons into focus at a point just in front of the anode, at which point they diverge, and are then brought to a second focal point by the fields in the region between the three anodes. A focus control is provided which by adjustment of the potential at the third anode makes the focal point coincide with the position of the screen. When the electrons impact on the screen coating, the phosphor material luminesces at the beam focal point, causing emission of a spot of light on the face of the screen.

In order to 'trace out' a luminescent display, it is necessary for the spot of light to be deflected about horizontal and vertical axes, and for this purpose a beam-deflection system is also provided. Deflection systems can be either electrostatic or electromagnetic, the latter being used in the tubes applied to the display units of aircraft systems.

The manner in which an electromagnetic field is able to deflect an electron beam is illustrated in Fig. 22.4. A moving electron constitutes an electric current, and so a magnetic field will exist around it in the same way as a field around a current-carrying conductor. In the same way that a conductor will experience a deflecting force when placed in a permanent magnetic field, so an electron beam can be forced to move when subjected to electromagnetic fields acting across the space within the tube. Coils are, therefore, provided around the neck of the tube, and are configured so that fields are produced horizontally (X-axis fields) and vertically (Y-axis fields). The coils are connected to the signal sources whose variables are to be displayed, and the electron beam can be deflected to the left or right, up or down, or along some resultant direction depending on the polarities produced by the coils, and on whether one alone is energized, or both are energized simultaneously.

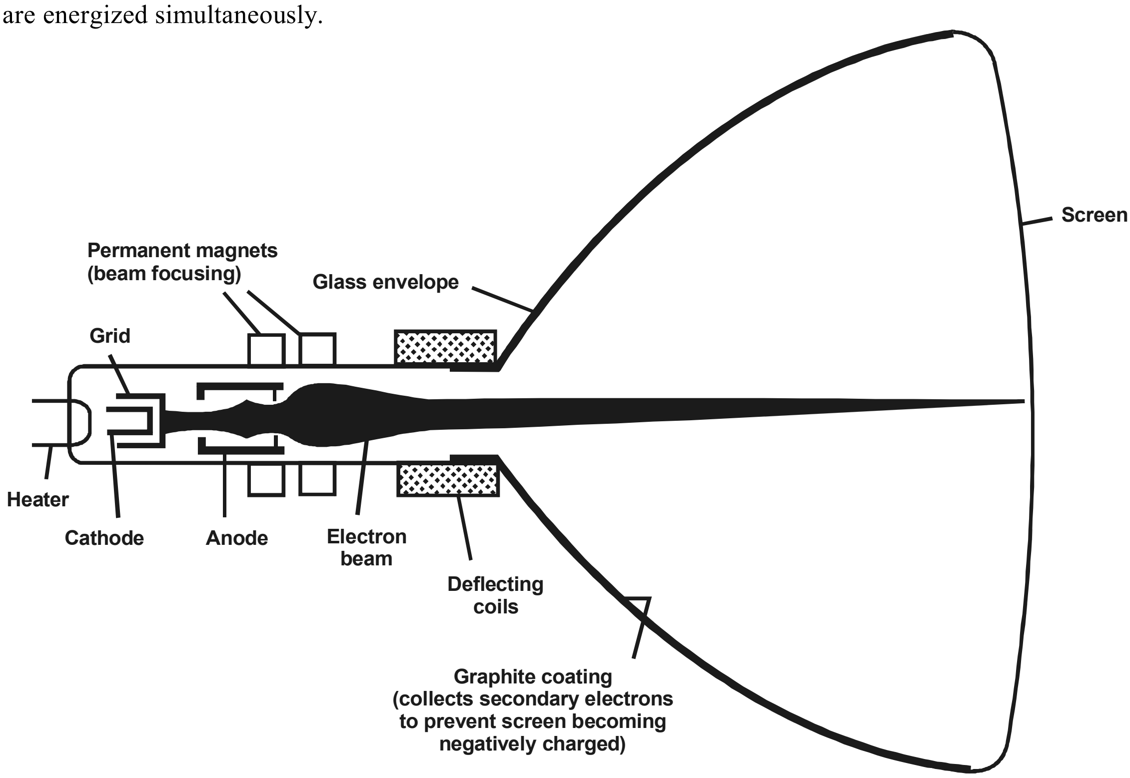



*Fig.22.3. Cathode ray tube.*

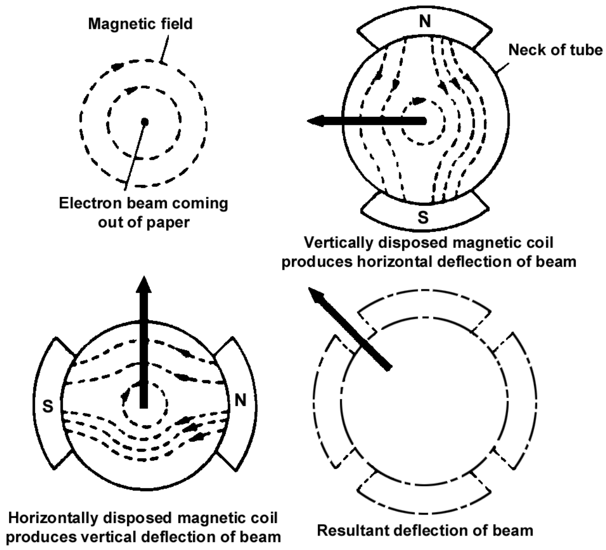



*Fig.22.4. Electron beam deflection.*

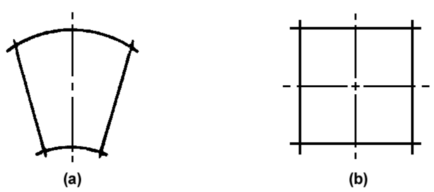



*Fig.22.5. Data cells. (a) Rho-theta; (b) X-Y coordinate.*

## COLOUR CRT DISPLAYS

These are used in weather radar display units, and are the norm for those units designed for the display of data associated with the systems installed in the types of aircraft referred to earlier. In these display units weather data is also integrated with the other data displays, and since there is a fundamental similarity between the methods through which they are implemented, the operation of a weather radar display unit serves as a useful basis for study of the display principles involved.

The video data received from a radar antenna is conventionally in what is termed rho-theta form, corresponding to the 'sweeping' movement of the antenna as it is driven by its motor [see Fig. 22.5(a)]. In a colour display indicator, the scanning of data is somewhat similar to that adopted in the tube of a television receiver, i.e. *raster* scanning in horizontal lines. The received data is still in rho-theta form, but in order for it to be displayed it must be converted into an X-Y coordinate format as shown in Fig. 22.5 (b). This format also permits the displayed. In addition it permits a doubling-up of the number of data cells, as indicated by the dotted lines in the diagram.

Each time the radar transmitter transmits a pulse, the receiver begins receiving return echoes from 'targets' at varying distances (rho) from the transmitter. This data is digitized to provide output levels in binary-coded form, and is supplied to the indicator on two corresponding to the level of the return echoes which, in turn, are related to the weather conditions prevailing at the range in nm preselected on the indicator. The data are stored in memories which, on being addressed as the CRT is scanned, will at the proper time permit the weather condition to be displayed. The four conditions are displayed as follows:

| | |
|---|---|
| Blank screen : | Zero or low-level returns. |
| Green : | Low returns (lowest rainfall rate). |
| Yellow : | Moderate returns (moderate rainfall rate). |
| Red : | Strong returns (high-density rainfall rate). |

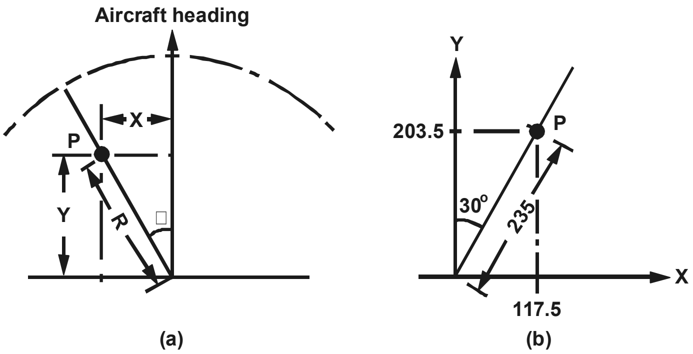



*Fig.22.6. Scan conversion.*

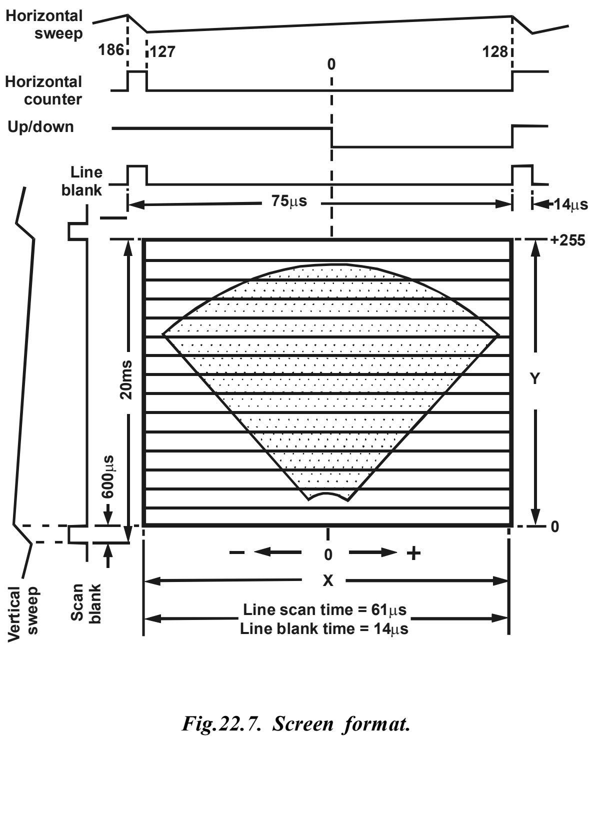



*Fig.22.7. Screen format.*

**Scan conversion**

The principle of conversion from rho-theta form to an X - Y coordinate scan is shown in Fig. 22.6. With a 'target' at point P, at a range R and an angle $\theta$, it will have coordinates: $X = R \sin \theta$ and $Y = R \cos \theta$ . Thus, for an echo received at an azimuth angle of, say, $30^o$ and a range of 235 nm, the coordinates will be: $X = 235 \sin 30^o = 117.5$ nm, and $Y = 235 \cos 30^o = 203.5$ nm. The conversion is performed by a microprocessor on the indicator's display circuit board.

**Screen format**

The coordinate system format of the screen is shown in Fig. 22.7, and from this example it will be noted that the screen is divided into two halves representing two quadrants in the coordinate system. The origin is at the bottom centre, so that values of X are negative to the left and positive to the right; all values of Y are positive. The screen is scanned in 256 horizontal lines, and there are 256 'bits' of information displayed on each line.

Each line is located by a value of Y and each bit by a value of X; the screen therefore has a 256X 256 matrix. The X and Y values are used to address the memory and display the information stored there as the appropriate time in the scan occurs. The memory for the weather data is in two parts which store the bits of the data words that represent the colours red, yellow or green and the corresponding weather conditions. Each part of the memory contains one address for every bit on every line in the display; each memory, therefore, is also a 256X256 matrix, and allows the entire weather display to be stored continuously.

As the screen is scanned, the memory is addressed at each point on each line by two counters: a horizontal or X counter for addressing the rows in the memory, and a vertical or Y counter for addressing the columns. The X counter generates an output for each of the 256 bits on a line, and counting is started by a 'high' state output signal from an up/down divider circuit. The counter is caused to count down, i.e. left to right, from the number 186 to 0 at the centre of the screen. When it reaches 0, the divider circuit changes to a 'low' state output, thereby causing the counter to count up to the number 128 at the end of the line, at which point a 'line blank' pulse of 14 μs duration is generated. The line scan time is about 61 μs, and so the total time for each line is 75 μs. The divider circuit again changes to a 'high' state to cause the counter to start down for the next line, and is a process that is repeated for all remaining lines.

An output from the X counter is also applied to the Y counter, which counts to 256 (one for each line) plus eight counts for a scan blank time to allow for the CRT beam 'spot' to return to the upper left corner of the screen. This process is repeated, and since there are 256 lines in the display it takes 20 ms to scan the entire screen (19.4 ms for the 256 lines and 600 μs for the scan blank time). The vertical and horizontal sweep circuits are synchronized by the triggering of the line and scan blank pulses.

In addition to the foregoing raster scanning technique, which produces sections of a CRT screen in 'solid' colour, a *stroke* scanning technique is also used for producing displays of symbols and of data in alphanumeric format. Details of this will be given later in this chapter.

**Colour generation**

A colour CRT has three electron guns, each of which can direct an electron beam at the screen which is coated with three different kinds of phosphor material. On being bombarded by electron beams, the phosphors luminesce in each of the three primary colours red, green and blue

The screen is divided into a large number of small areas or dots, each of which contains a phosphor of each kind as shown in Fig. 22.8. The beam from a particular gun must only be able to strike screen elements of one colour, and to achieve this a perforated steel sheet called a *shadow mask* is accurately positioned adjacent to the coating of the screen. The perforations are arranged in a regular pattern, and their number depends on the size of screen; 330 000 is typical.

Beams emitted from each gun pass through the perforations in the mask and they cause the phosphor dots in the coating to luminesce in the appropriate colour. For example, if a beam is being emitted by the 'red' electron gun only, then only the red dots will luminesce, and if the beam completes a full raster scan of the screen, then as a result of persistence of vision by the human eye, a completely red screen will be observed. In the display units of electronic instrument systems, a number of other colours are also required and these are derived by independent circuit control of the three guns and their beam currents, so that as the beams strike the corresponding phosphor dots, the basic process of mixing of primary colours takes place (see Fig. 22.9). In other words, an electronic form of 'paint mixing' is carried out.

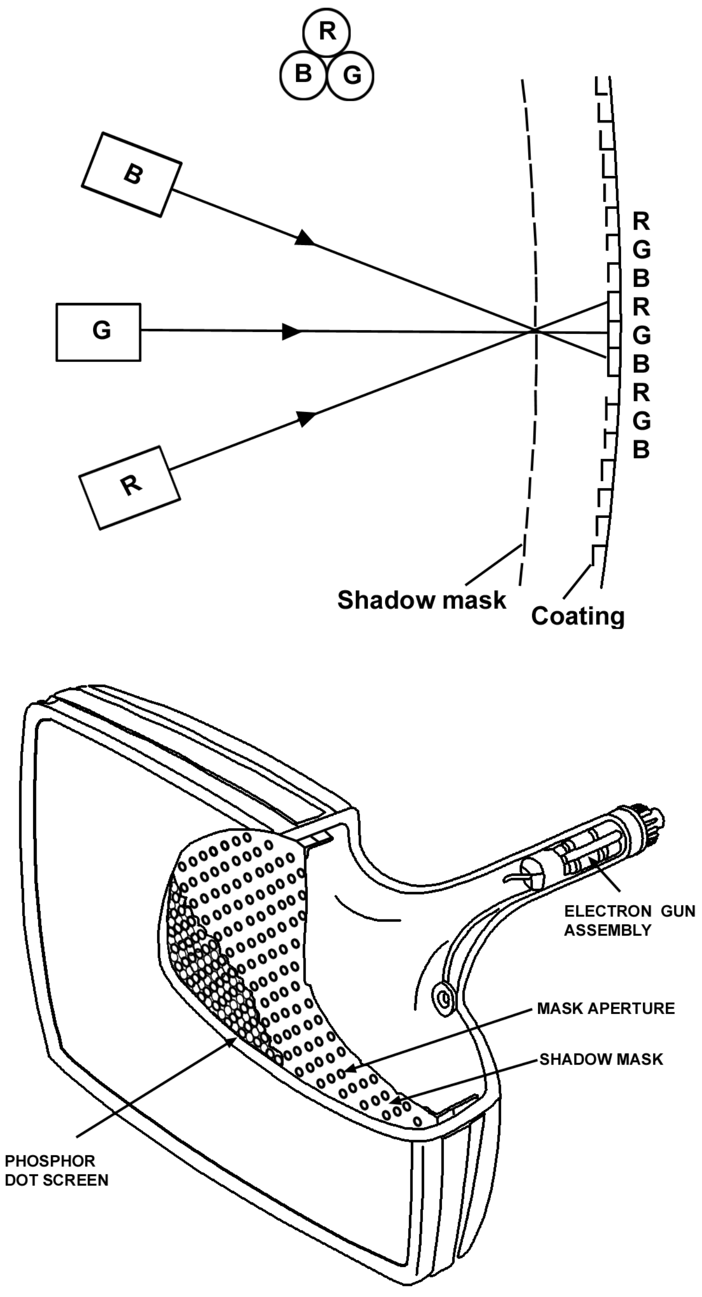



*Fig.22.8. Colour CRT.*

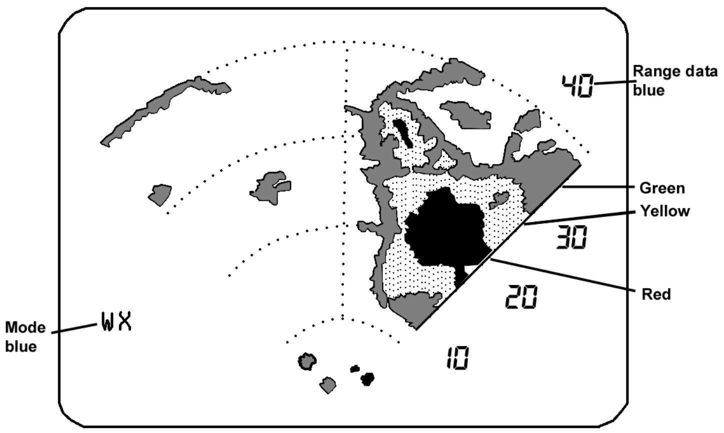



*Fig.22.9. Weather data display.*

Referring once again to the weather radar indicator application, the data readout from the memory, apart from being presented at the appropriate location of the CRT screen, must also be displayed in the colours corresponding to the weather conditions prevailing. In order to achieve this, the data is decoded to produce outputs which, after amplification, will turn on the requisite colour guns; the data flow is shown in Fig. 22.10. The memory output is applied

**Table 22.1**

| Inputs | | Outputs | | | Colour |
|---|---|---|---|---|---|
| M | L | $A_1$ | $A_2$ | $A_3$ | |
| 0 | 0 | 1 | 1 | 1 | Black (off) |
| 0 | 1 | 0 | 1 | 1 | Green |
| 1 | 0 | 1 | 0 | 1 | Yellow |
| 1 | 1 | 1 | 1 | 0 | Red |

**Table 22.2**

| Outputs to guns | | | Resulting colours |
|---|---|---|---|
| $B_1$ Green | $B_0$ Blue | $B_2$ Red | |
| 1 | 1 | 1 | Black (off) |
| 0 | 0 | 0 | White |
| 0 | 0 | 1 | Yellow |
| 0 | 1 | 1 | Red |
| 1 | 0 | 0 | Light Blue |
| 1 | 0 | 1 | Green |

to a data demultiplexer whose output corresponds to the most significant and least significant bits (M and L) of the two-bit binary words and is supplied to a data decoder. The inputs are decoded to provide three-bit output words corresponding to the colours to be displayed, as shown in Table 22.1. The outputs are then applied to the colour decoder and primary encoder circuit, and this in turn provides three outputs, each of which corresponds to one of the colour guns as shown in Table 22.2.

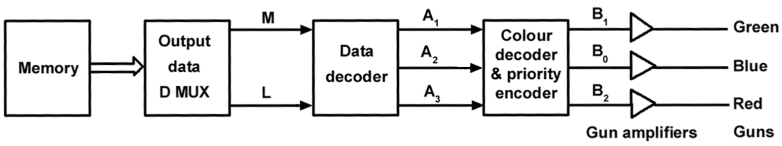



*Fig.22.10. Data flow for gun operation.*

The 'low' state outputs turn on the guns, and from Table 22.2 it can also be seen how simultaneous gun operation produces other colours from a mix of the primary colours. Figure 22.9 illustrates a typical weather data display together with associated alphanumeric data, namely ranges in nm, and an operating mode which in this case is WX signifiying 'weather' mode.

**ALPHANUMERIC DISPLAYS**

The display of data in alphanumeric and in symbolic form is extremely wide-ranging. For example, in a weather radar indicator it is usually only required for range information and indications of selected operating modes to be displayed, while in systems designed to perform functions within the realm of flight management, a very much higher proportion of information must be 'written' on the screens of the relevant display units. This is accomplished in a manner similar to that adopted for the display of weather data, but additional memory circuits, decoders, and character and symbol generator circuits are required.

Raster scanning is also used, but where datum marks, arcs or other cursive symbols are to be displayed, a *stroke pulse* method of scanning is adopted. The position of each character on the screen is predetermined and stored in a memory matrix, typically 5 × 7, and when the matrix is addressed, the character is formed within a corresponding matrix of dots on the screen by video signal pulses produced as the lines are scanned.

Figure 22.11 illustrates how, for example, the letters 'WX' and the number '40' are formed. One line of dots is written at a time for the area in which the characters are to be displayed, and so for a 5X 7 matrix, seven image lines are needed to write complete characters and/or row of characters. As will be noted from Fig. 22.9, the characters are displayed in blue, so only the 'blue' electron gun is active in producing them. Spacing is necessary between individual characters and also between rows of characters, and so extra line 'blanking bits', e.g. three, are allocated to character display areas.

In the example of the weather radar indicator, the characters each have an allocation of eight bits (five for the characters and three for the space following) on each of 21 lines (14 for the character and seven for the space below). The increase in character depth to 14 lines is derived from an alphanumeric address generator output that writes each line in a character twice during line scanning. The character format in this case permits the display of 12 rows each of 32 characters.

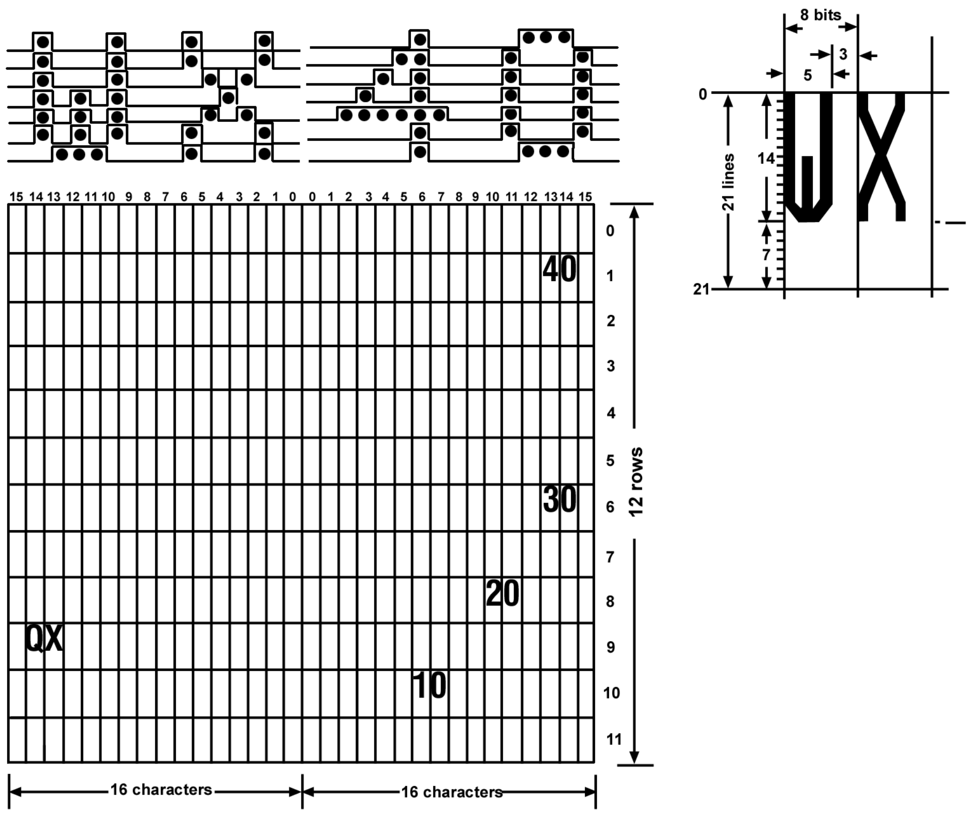



*Fig.22.11. Alphanumeric display.*

The CRT display units of the more comprehensive electronic instrument systems (see Chapters 12 and 16) operate on the same fundamental principles as those described, but in applying them, more extensive microprocessing circuit arrangements are required in order to display far greater amounts of changing data in quantitative and qualitative form.

The microprocessor processes information from the data 'highway' bus and, from the memory circuits, it is instructed to call up sub-programs, each of which corresponds to the individual sets of data that are required to be displayed. Signals are then generated in the relevant binary format, and are supplied to a symbol generator unit. This unit, in turn, generates and supplies signals to the beam deflection and colour gun circuits of the CRT, such that its beams are raster and stroke scanned, to present the data at the relevant parts of the screen, and in the required colour.

The displayed data is in two basic forms: fixed and moving. Fixed data relate in particular to such presentations as symbols, scale markings, names of systems, datum marks, names of parameters being measured, etc. Moving data are in the majority, of course, since they present changes occurring in the measurement of all parameters essential for in-flight management. The changes are indicated by the movement of symbolic pointers, index marks, digital counter presentations, and system status messages, to name a few.

■■■

# ELECTRONICS/ RADIO SYSTEM

# CHAPTER : 23
# SEMICONDUCTOR DIODES

## P.N JUNCTION DIODE

### (a) Construction

It is a two-terminal device consisting of a P-N junction formed either in Ge or Si crystal. Its circuit symbol is shown in Fig. 23.1 (a) The P-type and N-type regions are referred to as anode and cathode respectively. In Fig.1 (b) arrowhead indicates the conventional direction of current flow when forward-biased. It is the same direction in which hole flow takes place.

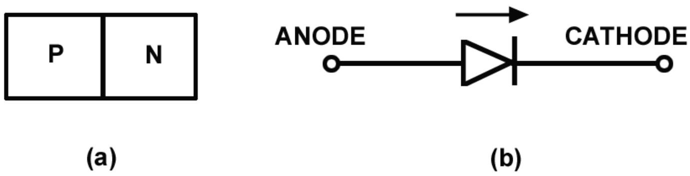



*Fig. 23.1.*

Commercially available diodes usually have some means to indicate which lead is P and which lead is N. Standard notation consists of type numbers preceded by 'IN' such a IN 240 and IN 1250. Here, 240 and 1250 correspond to colour bands.

Fig. 23.2 (a) shows typical diodes having a variety of physical structures whereas Fig.23.2 (b) shows terminal identifications.

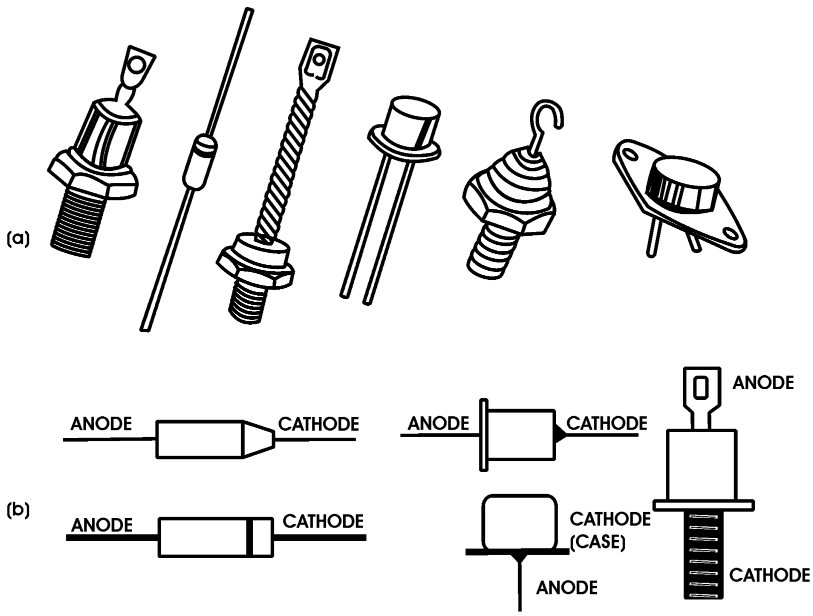



*Fig. 23.2.*

### (b) Working

A P-N junction diode is a one-way device offering low resistance when forward-biased and behaving almost as an insulator when reverse-biased. Hence, such diodes are mostly used as rectifiers i.e. for converting alternating current into direct current.

### (c) V/I Characteristics

Fig. 23.3 shows the static voltage-current characteristics for a low-power P-N junction diode.

### *(i) Forward Characteristic*

When the diode is forward-biased and the applied voltage is increased from zero, hardly any current flows through the device in the beginning. It is so because the external voltage is being opposed by the internal barrier voltage $V_B$ whose value is 0.7 V for Si and 0.3 V for Ge. As soon as $V_B$ is neutralised, current through the diode increases rapidly with increasing battery voltage. It is found that as little a voltage as 1.0 V produces forward current of about 50 mA.

***(ii) Reverse Characteristic***

When the diode is reverse-biased, majority carriers are blocked and only a small current (due to minority carriers) flows through the diode. As the reverse voltage is increased from zero, the reverse current very quickly reaches its maximum or saturation value $I_o$ which is also known as leakage current. It is of the order of nanoamperes (nA) for Si and microamperes (mA) for Ge.

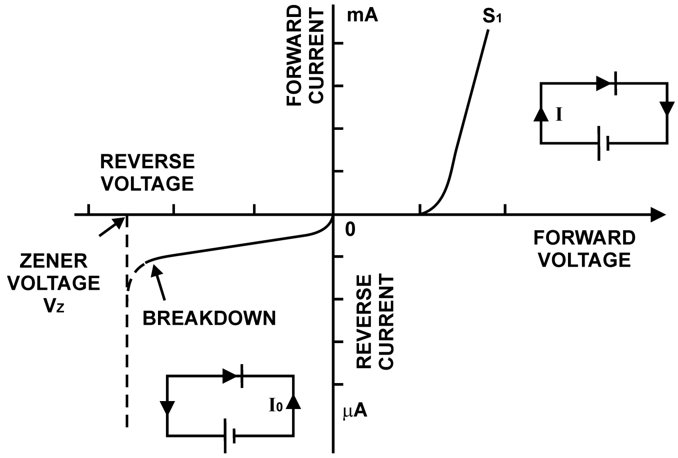



*Fig. 23.3.*

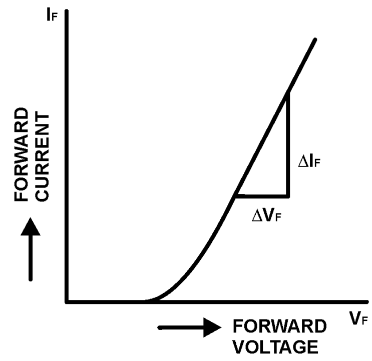



*Fig. 23.4.*

As seen from Fig. 23.3 When reverse voltage exceeds a certain value called Zener voltage $V_Z$, the leakage current suddenly and sharply increases, the curve indicates zero resistance at this point. When P-N junction diodes are employed primarily because of this property as voltage regulators, they are called Zener diodes.

**(d) Diode Parameters**

The diode parameters of greatest interest are :

***1. Bulk resistance ($r_B$)***

It is the sum of the resistance values of the P- type and N- type semiconductor materials of which the diode is made of. Usually, it is very small.

***2. Junction resistance ($r_j$)***

Its value for forward-biased junction depends on the magnitude of forward dc current.

$r_j \cong 25\,\text{mV}/I_F$ .......for Ge

$\cong 50\,\text{mV}/I_F$ .......for Si

Where $I_F$ is the forward current in mA.

***3. Dynamic or ac resistance***

$$r_{ac} \text{ or } r_d = r_B + r_j$$

As seen from forward characteristic of Fig. 23.4

$r_{ac} = \Delta V_F / \Delta I_F$ .....at a given dc forward current ***Forward voltage drop***

It is given by the relation :

***4. Forward voltage drop*** $= \dfrac{\text{power dissipated}}{\text{forward dc current}}$

***5. Reverse saturation current ($I_o$)***

It has already been discussed.

***6. Reverse breakdown voltage ($V_{BR}$)***

It has already been discussed.

***7. Reverse dc resistance $R_R$***

$$R_R = \frac{\text{reverse voltage}}{\text{reverse current}}$$

**(e) Applications**

The main applications of semiconductor diodes in modern electronic circuitry are as under :

1. *As rectifier diodes.* They convert a.c. current into d.c. current for d.c. power supplies of electronic circuits.
2. As signal diodes in communication circuits for modulation and demodulation of small signals.
3. As Zener diodes in voltage stabilizing circuits.
4. As varactor diodes-for use in voltage controlled tuning circuits as may be found in radio and TV receivers. For this purpose, the diode is deliberately made to have a certain range of junction capacitance. The capacitance of the reversed-biased diode is given by $C = K/\sqrt{V}$.
5. In logic circuits used in computers.

**DIODE AS A RECTIFIER**

Conversion of alternating current (or voltage) into direct current (or voltage) is called rectification. A diode is well-suited for this job because it conducts only in one direction i.e. only when forward-biased. Half-wave rectifier uses one diode, full -wave rectifier uses two diodes whereas a bridge rectifier uses four diodes.

In India, electric energy is available at an alternating voltage of 220 V (r.m.s. value) at 50 Hz. However, for the operation of most of the electronic equipment, dc voltage is needed. Hence, practically, every electronic equipment includes a circuit which provides it with a stabilized dc output voltage from the input ac voltage. In fact, this circuit forms the power supply of the equipment and consists of :

1. A transformer - for either stepping up or down the ac supply voltage.
2. Rectifier - for converting this stepped up (or down) ac voltage into dc voltage.
3. Smoothing filter - for removing any variations or ripples in the dc output voltage.
4. Stabilizer - for keeping output dc voltage constant even when input voltage or load current changes.

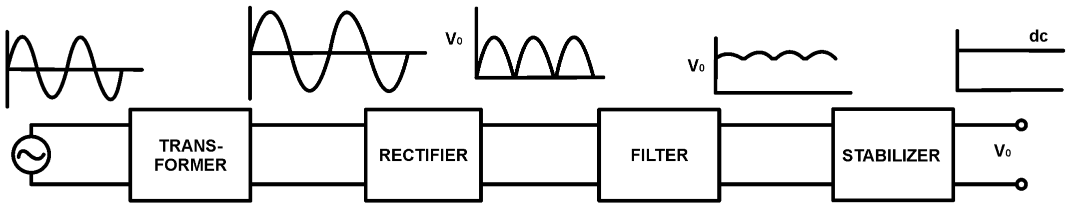



*Fig. 23.5.*

A block diagram of such a power supply is shown in Fig. 23.5 Obviously, rectifier is the heart of a power supply.

Incidentally, in Fig. 23.5 full - wave rectifier has been assumed.

**HALF-WAVE RECTIFIER**

Though such a circuit is not much used, yet we will discuss it here for the sake of explaining the basic principle involved in the working of a rectifier. The simple circuit of a half-wave rectifier is shown in Fig. 23.6 along with the input and output voltage wave-forms. An ac voltage is applied to a single diode connected in series with a load resistor $R_L$.

**a) Working**

During the positive half-cycle of the input ac voltage i.e. when point M is positive, the diode D is forward-biased (ON) and conducts. While conducting, the diode acts as a short-circuit so that circuit current flows and hence, positive half-cycle of the input ac voltage $v_o$ as shown in Fig.23.6 (b). Waveform of the diode current (which equals load current) is similar to the voltage waveform.

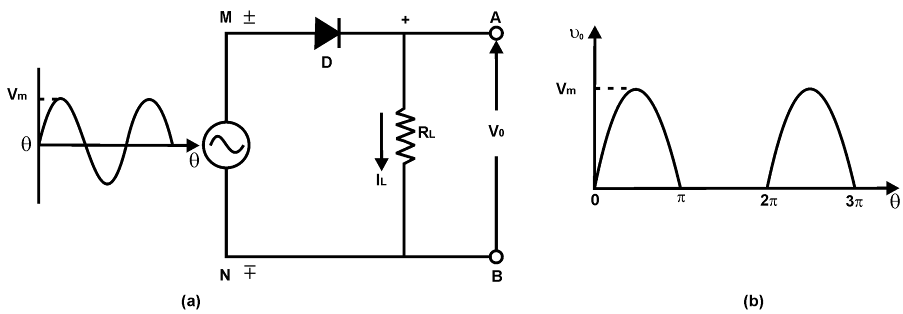



*Fig. 23.6.*

During the negative input half-cycle i.e. when point M becomes negative, the diode is reverse-biased ( OFF ) and, so, does not conduct i.e. there is no current flow. Hence, there is no voltage drop across $R_L$. In other words $i_L = 0$ and $v_L = v_o = 0$. Obviously, the negative input half-cycle is suppressed i.e. it is not utilized for delivering power to the load. As seen, the output is not a steady dc but only a pulsating dc wave having a ripple frequency equal to that of the input voltage frequency. This wave can be observed by an oscilloscope connected across $R_L$. When measured by a dc meter, it will show some average positive value both for voltage and current. Since only one half-cycle of the input wave is used, it is called a half-wave rectifier. It should be noted that forward voltage drop across the diode has been neglected in the above discussion. We have, in fact, assumed an ideal diode. If it is required to step up or step down the input voltage, we will have to use a power transformer as shown in Fig. 23.7

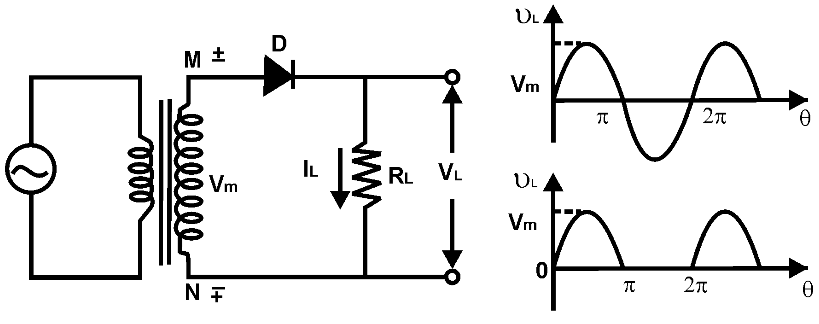



*Fig. 23.7.*

Another advantage of using this transformer is that it isolates the circuit from the ac mains thereby reducing the risk of electrical shock.

**b) Average Values**

The output voltage of a half-wave rectifier is found to consist of positive half-cycles only. Their average value is (Fig. 23.8)

$$V_{av} = V_{dc} = \frac{Vm}{\pi} = 0.318\ V_m$$

Similarly,

$$I_{av} = I_{dc} = \frac{Im}{\pi} = 0.318\ I_m$$

$V_{dc}$ is the dc component of the load voltage $V_L$ and, in a similar way, $I_{dc}$ is the dc component of $I_L$. Both $V_L$ and $I_L$ have ac components.

**FULL-WAVE RECTIFIER**

In this case, both half-cycles of the input are utilized with the help of two diodes working alternately. For full-wave rectification, use of a centre-tap transformer is essential (though it is optional for half-wave rectification).

The full-wave rectifier circuit using two diodes and a centre-tapped transformer is shown in Fig. 23.9 The centre-tap is usually taken as the ground or zero voltage reference point.

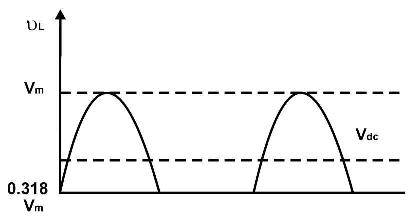



*Fig. 23.8.*

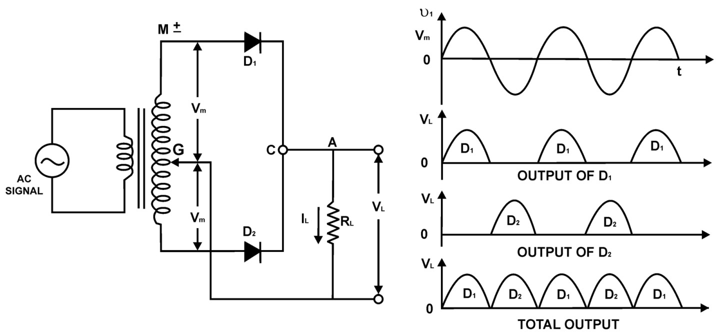



*Fig. 23.9.*

Fig. 23.10 shows two different ways of drawing the circuit. In Fig. 23.9 $R_L$ becomes connected to point G via the earth whereas in Fig. 23.10 , it is connected directly to G.

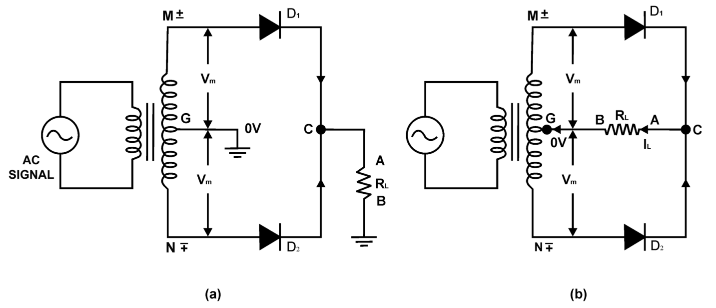



*Fig. 23.10.*

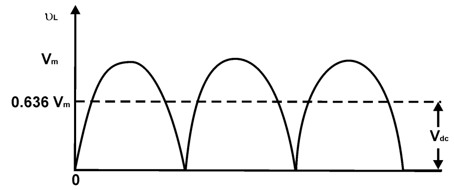



*Fig. 23.11.*

### (a) Working

When input ac supply is switched on, the ends M and N of the transformer secondary becomes positive and negative alternately. During the positive half-cycle of the ac input, terminal M is positive, G is at zero potential and N is at negative potential. Hence, being forward-biased diode $D_1$ conducts ( but not $D_2$ which is reverse-biased) and current flows along $MD_1$ CABG. As a result, positive half-cycle of the voltage appears across $R_L$.

During the negative half-cycle, when terminal N becomes positive, then $D_2$ conducts (but not $D_1$ ) and current flow along $ND_2$ CABG. So, we find that current keep on flowing through $R_L$ in the same direction (i.e. from A to B) in both half-cycles of the ac input. It means that both half-cycles of the input ac supply are utilized as shown in Fig, 23.9 (b) the frequency of the rectified output voltage is twice the supply frequency.

### (b) Average Values

As seen from Fig. 23.11 the output voltage of a full-wave rectifier consists of both positive and negative half-cycles. The dc rectifier consists of both positive and negative half-cycles. The dc value of load voltage $V_L$ is given by

$V_{dc} = (2\,V_m)/\pi = 0.636\,V_m$

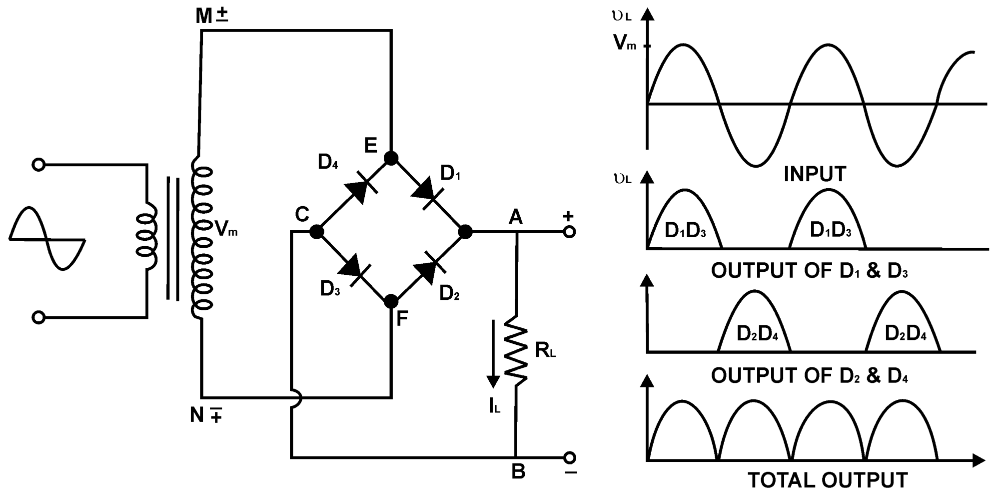
= twice the HW rectified value

Where $V_m$ is the maximum voltage across each half of the transformer secondary. Similarly, the dc component of load current is

$$I_{dc} = (2\,I_m) / \pi = 0.636\,I_m = 0.636\,(Vm)/R_L$$

**FULL-WAVE BRIDGE RECTIFIER**

It is the most frequently-used circuit for electronic dc power supplies. It requires four diodes but the transformer used is not centre-tapped and has a maximum voltage of $V_m$ across its secondary. The circuit using four discrete diodes is shown in Fig. 23.12.




(a) (b)

*Fig. 23.12*

### (a) Working

During the positive input half-cycle, terminal M of the secondary is positive and N is negative as shown separately in Fig. 23.13 (a) Diodes $D_1$ and $D_3$ becomes forward-biased (ON) whereas $D_2$ and $D_4$ are reverse-biased (OFF). Hence, current flows along the path MEABCFN producing a drop across $R_L$.

During the negative input half-cycle, secondary terminal N becomes positive and M negative. Now, $D_2$ and $D_4$ are forward-biased. Circuit current flows along NFABCEM as shown in Fig. 23.13 (b) Hence, we find that current keeps flowing through load resistance $R_L$ in the same direction AB during both half-cycles of the ac input supply. Consequently, point A of the bridge rectifier always acts as an anode and point B as cathode.

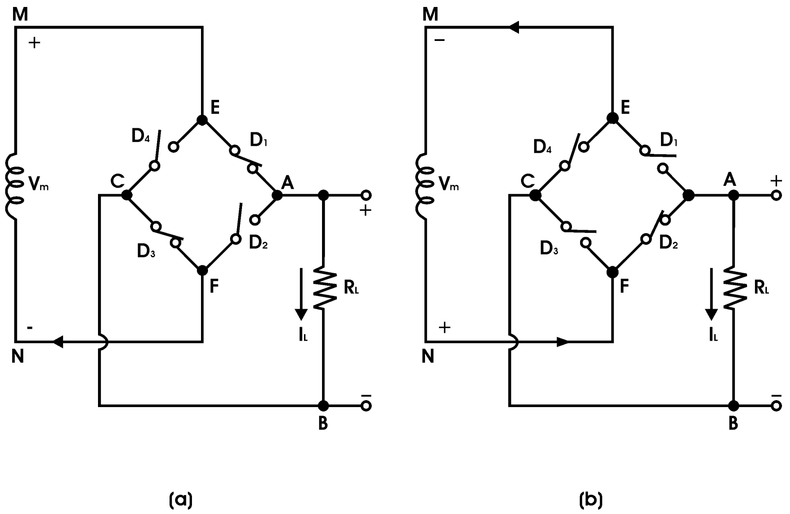



*Fig. 23.13*

The output voltage across $R_L$ is as shown in Fig. 23.12 (b) Its frequency is twice that of the supply frequency. Moreover, for a given power transformer, a bridge rectifier produces an output voltage nearly twice as large as produced by an ordinary full wave rectifier. It is due to the fact that the entire secondary voltage of the transformer is applied across each of the diode pairs.

**(b) Average Values**

$V_{dc} = 0.636V_m$ and $I_{dc} = 0.636/3\ I_m$ here $I_m = V_m/R_L$

**(c) Advantages**

After the advent of low cost, highly reliable and small sized silicon diodes, bridge circuits has become much more popular than the centre-tapped transformer Full Wave Rectifier. The main reason for this is that for a bridge rectifier, a much smaller transformer is required for the same output because it utilizes the transformer secondary continuously unlike the 2- diode FW rectifier which uses the two halves of the secondary alternately.

So, the advantages of the bridge rectifier are

1. No centre-tap is required on transformer
2. Much smaller transformers are required
3. It is suitable for high-voltage applications
4. It has less PIV rating per diode.

The obvious disadvantage is the need for twice as many diodes as for the centre-tapped transformer version. But ready availability of low - cost silicon diodes has made it more economical despite its requirement of four diodes.

**ZENER DIODE**

It is a reverse-biased heavily-doped silicon or germanium P-N junction diode which is always reverse-biased and operates in the breakdown region where current is limited only by both external resistance and the power dissipation of the diode. Conventional diodes and rectifiers never operate in the breakdown region but the Zener diode makes a virtue of it and operates at this very point like P shown in Fig. 23.14 The Zener breakdown occurs due to breaking of covalent bonds by the strong electric field set up in the depletion region by the reverse voltage.

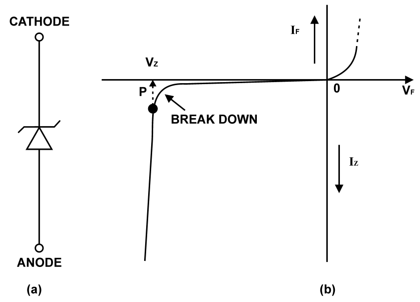



*Fig. 23.14*

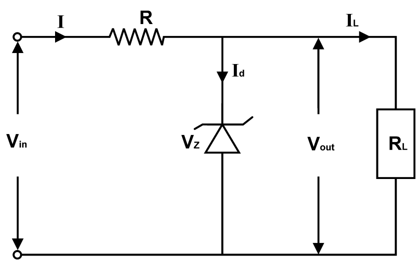



*Fig. 23.15*

Fig. 23.14 (a), shows schematic symbol of a Zener diode and is similar to that of a normal diode except that the line representing the cathode is bent at both ends. With a little mental effort, the cathode symbol can be imagined to look like the letter Z for Zener.

The typical V/I characteristic is shown in Fig. 23.14 (b).

**Zener Diode As Voltage Regulator**

Voltage regulation is a measure of a circuit's ability to maintain a constant output voltage even when either input voltage or load current varies. A Zener diode when working in the breakdown region can serve as a voltage regulator. In Fig. 23.15 $V_{in}$ is the input dc voltage whose Variations are to be regulated. The zener diode is reverse-connected across $Z_{in}$. When potential difference across the diode is greater than $V_z$, it conducts and draws relatively large current through the resistance R. The load resistance $R_L$ across which a constant voltage $V_{out}$ is required, is connected in parallel with the diode. The total current I passing through R equals the sum of diode current and load current i.e. $I = I_d + I_L$ .

It will be seen that under all conditions, $V_{out} = V_z$.

Hence, $V_{in} = IR + V_{out} = IR + V_z$

**Case 1**

Suppose R is kept fixed but supply voltage $V_{in}$ is increased slightly. It will increase I. This increase in I will be absorbed by the Zener diode without affecting $I_L$. The increase in $V_{in}$ will be dropped across R thereby keeping $V_{out}$ constant. Hence , when $V_{in}$ changes, I and IR drop change is such a way as to keep $V_{out}$ ( $= V_z$ ) constant.

**Case 2**

In this case, $V_{in}$ is fixed but $I_L$ is changed. When $I_L$ increases, diode current $I_d$ decreases thereby keeping I and IR drop constant. In this way, $V_{out}$ remains unaffected.

Should $I_L$ decrease, $I_d$ would increase in order to keep I and hence IR drop constant.

Again, $V_{out}$ would remain unchanged because

$$V_{out} = V_{in} - IR = V_{in} - (I_d + I_L) R$$

Incidentally, it may be noted that

$$R = [( V_{in} - V_{out} )/ (I_d + I_L)]$$

It may also be noted that when diode current reaches its maximum value, $I_L$ becomes zero. In that case

$$R = [(V_{in} - V_{out} ) / I_{d(max)}]$$

**Zener Diode For Meter protection**

Zener diode are frequently use in volt-ohm-milli ammeters for protecting meter movement against burnout from accidental overloads. If VOM is set to its 2.5 - V range and the test leads are accidentally connected to a 25- V circuit, an unprotected meter will be burned out or at least get severely damaged.

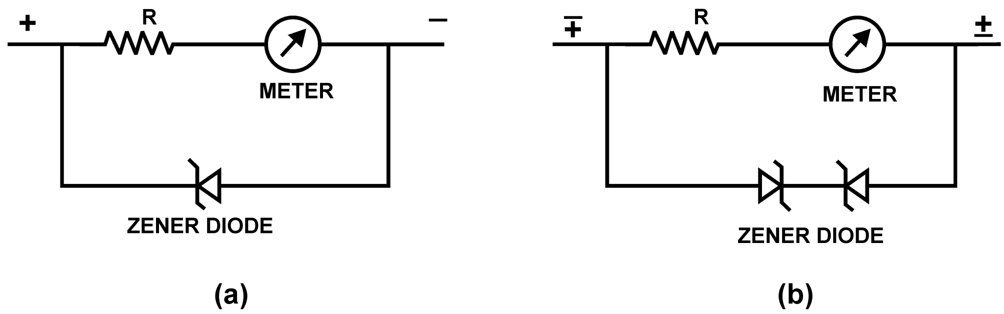



*Fig. 23.16*

This hazard can be avoided by connecting a Zener diode in parallel with the meter as shown in Fig. 23.16 (a) In the event of an accidental overload, most of the current will pass through the diode. Two Zener diodes connected as shown in Fig. 23.16 (b) can provide overload protection regardless of the applied polarity.

**Zener Diode as Peak Clipper**

Use of Zener diodes in wave-shaping circuits is illustrated in Fig. 23.17. Here, a semi-square wave is produced by clipping a sine wave. Two Zener diodes are shunted across a sine wave source. The shunt resistance remains very high until the positive and negative voltage across the diodes exceeds the Zener value. Thereafter, the diodes become low-resistance shunt circuits and increasing voltage values are dropped across the series resistance R. Consequently, the output wave is clipped on both peaks.

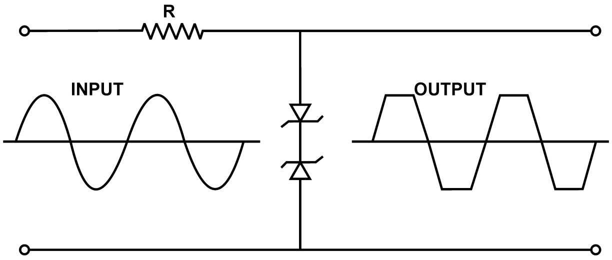



*Fig. 23.17*

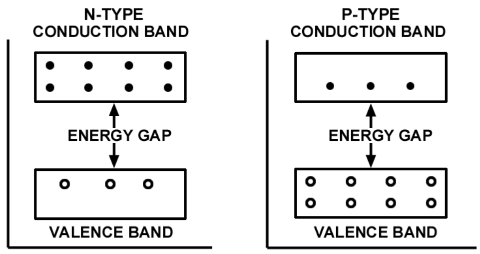



*Fig. 23.18*

**TUNNELING EFFECT**

In a normally-doped P-N junction, the depletion layer is relatively wide and a potential barrier exists across the junction. The charge carriers on either side of the junction cannot cross over unless they possess sufficient energy to overcome this barrier (0.3 V for Ge and 0.7 V for Si). As is well-known, width of the depletion region depends directly on the doping density of the semiconductor. If a P-N junction is doped very heavily (about 1000 times or more), its depletion layer becomes extremely thin (about 0.00001 mm). It is found that under such conditions, may carriers can 'punch through' the junction with the speed of light even when they do not posses enough energy to overcome the potential barrier. Consequently, large forward current is produced even when the applied bias is much less than 0.3 V.

This conduction mechanism in which charge carriers (possessing very little energy) bore through a barrier directly instead of climbing over it is called tunneling.

**Explanation**

Energy band diagrams (EBD) of N-type and P-type semiconductor of materials can be used to explain this tunneling phenomenon.

Fig. 23.18 shows they energy band diagram of the two types of silicon separately. As explained earlier , in the N-type semiconductor, there is increased concentration of electrons in the conduction band. It would be further increased under heavy doping. Similarly, in a P-type material, there is increased concentration of holes in the valence band for similar reasons.

**(a) No. Forward Bias**

When the N-type and P-type materials are joined, the Energy Band Diagrams under no-bias condition becomes as shown in Fig. 23.19 (a). The junction barrier produces only a rough alignment of the two materials and their respective valence and conduction bands. As seen, the depletion region between the two is extremely narrow due to very heavy doping on both sides of the junction. The potential hill is also increased as shown.

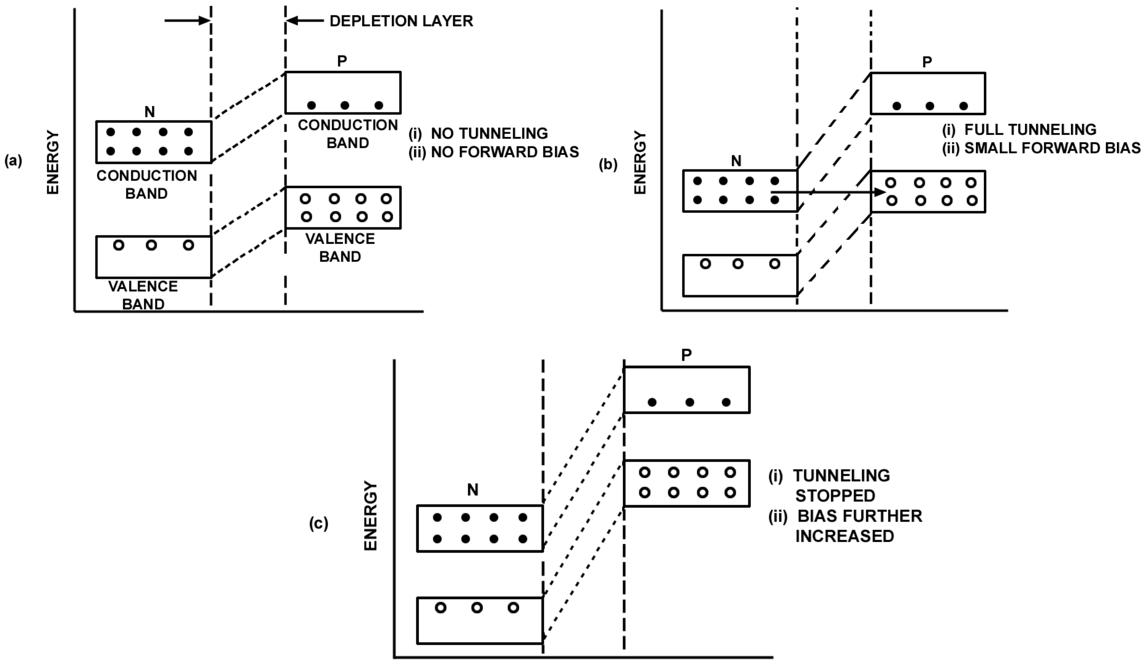



*Fig. 23.19*

**(b) Small Forward Bias**

When a very small forward voltage ($\simeq$ 0.1 V) is applied, the EBDs become as shown in Fig. 23.19 (b). Due to the downward movement of the N-region, the P-region valence band becomes exactly aligned with the N-region conduction band. At this stage, electrons tunnel through the thin depletion layer with the velocity of light thereby giving rise to a large current called peak current $I_P$.

**(c) Large Forward Bias**
When the forward bias is increased further, the two bands get out of alignment as shown in Fig 23.19 (c). Hence, tunneling of electrons stops thereby decreasing the current. Since current decreases with increase in applied voltage (i.e., dV/dl is negative), the junction is said to possess negative resistance at this stage. This resistance increases throughout the negative region.

However, it is found that when applied forward voltage is increased still further, the current starts increasing once again as in a normal junction diode.

**TUNNEL DIODE**
This diode was first introduced by Dr. Leo Easki in 1958.

**(a) Construction**
It is a high-conductivity two-terminal P-N junction diode having doping density about 1000 times higher as compared to an ordinary junction diode. This heavy doping produces following three unusual effects :

1. Firstly, it reduces the width of the depletion layer to an extremely small value (about 0.00001 mm).
2. Secondly, it reduces the reverse breakdown voltage to a very small value (approaching zero) with the result that the diode appears to be broken down for any reverse voltage.
3. Thirdly, it produces a negative resistance section on the V/I characteristic of the diode.

It is called a tunnel diode because due to its extremely thin depletion layer, electrons are able to tunnel through the potential barrier at relatively low forward bias voltage (less then 0.05 V). Such diodes are usually fabricated from germanium, gallium-arsenide (GaAs) and gallium antimonide (GaSb).

The commonly-used schematic symbols for the diode are shown in Fig. 23.20. It should be handled with caution because being a low power device, it can be easily damaged by heat and static electricity.

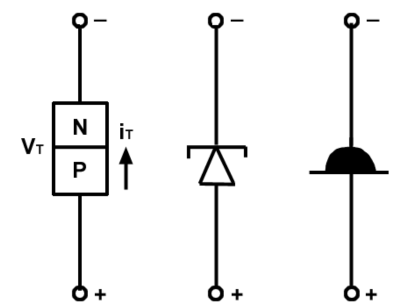



*Fig. 23.20*

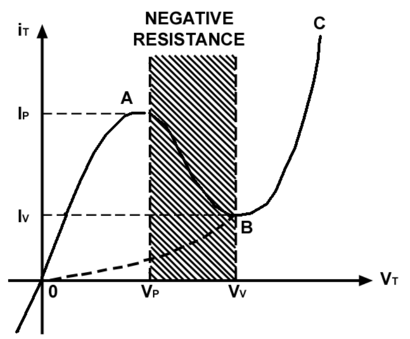



*Fig. 23.21*

**(b) V/I Characteristic**
It is shown in Fig. 23.21. As seen, forward bias produces immediate conduction i.e., as soon as forward bias is applied, significant current is produced. The current quickly rises to its peak value $I_P$ when the applied forward voltage reaches a value $V_P$ (point A). For voltages greater than $V_V$, current starts increasing again as in any ordinary junction diode.

As seen from Fig. 23.21, between the peak point A and valley point B, current decreases with increase in the applied voltage. In other words, tunnel diode possesses negative resistance ($-R_N$) in this region. In fact, this constitutes the most useful property of the diode. Instead of absorbing power, a negative resistance produces power. By offsetting losses in L and C components of a tank circuit, such a negative resistance permits oscillations. Hence, a tunnel diode is used as a very high frequency oscillator.

Another point worth noting is that this resistance increases as we go from point A to B because as voltage is increased, current keeps decreasing which means that diode negative resistance keeps increasing.

**(c) Tunneling Theory**
At zero forward bias, the energy levels of conduction electrons in N-region of the junction are slightly out of alignment with the energy levels of holes in the P-region. As the forward voltage is slightly increased, electron levels start getting aligned with the hole levels on the other side of the junction thus permitting some electrons to cross over. This kind of junction crossing is called tunneling.

As voltage is increased to peak voltage ($V_P$), all conduction band electrons in the N-region are able to cross over to the valence band in the P-region because the two bands are exactly aligned. Hence, maximum current (called peak current $I_P$) flows in the circuit.

After $V_P$, as the applied voltage is increased, current starts decreasing because the two bands start gradually getting out of alignment. It reaches minimum value (called valley current) when the two are totally out of alignment at a forward bias of $V_V$ (valley voltage).

For voltages greater than $V_V$, current starts increasing again exactly as it does in the case of an ordinary P-N junction diode.

Tunneling is much faster than normal crossing which enables a tunnel diode to switch ON and OFF much faster than an ordinary diode. That is why tunnel diode is extensively used in special application requiring very fast switching speeds like high-speed computer memories and high-frequency oscillators etc.

**(d) Diode Parameters**

***1. Negative Resistance ($-R_N$)***

It is the resistance offered by the diode within the negative-resistance section of its characteristic (shown shaded in Fig. 23.21). It equals the reciprocal of the slope of the characteristic in this region.

It may also be found from the following relation :

$$R_N = -\frac{dV}{dI}$$

Its value depends on the semiconductor material used (varying from -10$\Omega$ to -200$\Omega$).

***2. $I_P / I_V$ Ratio***

It is almost as important factor (particularly for computer applications) as the negative resistance of the diode.

Silicon diodes have a low $I_P / I_V$ ratio of 3:1 and their negative resistance can be approximated from $R_N = -220 / I_P$. Such diodes are used mainly for switches operating in high ambient temperatures.

Germanium diodes have an $I_P / I_V$ ratio of 6 : 1 and negative resistance formula $RN = -120 / I_P$.

GaAs diodes (used exclusively in oscillators) have an $I_P / I_V$ ratio of about 10 : 1 and a negative resistance nearly equal to that of silicon diodes.

The minimum $I_P / I_V$ ratio for GaSb diode is about 12 : 1 and has the lowest resistance of all given by $R_N = -60 / I_P$. Hence, such diodes have the lowest noise.

**(e) Equivalent Circuit**

The equivalent circuit of a tunnel diode is shown in Fig. 23.22. The capacitance C is the junction diffusion capacitance (1 to 10 pF) and ($-R_N$) is the negative resistance. The inductor $L_S$ is mainly due to the terminal leads (0.1 to 4 nH). The resistance $R_S$ is due to the leads, ohmic contact and semiconductor materials themselves (1 - 5$\Omega$). These factors limit the frequency at which the diodes may be used. They are also important in determining the switching-speed limit.

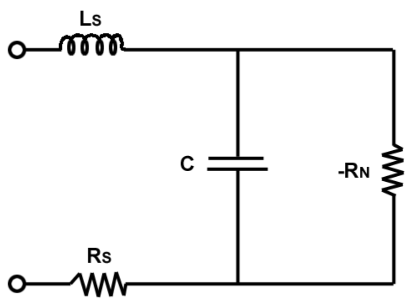
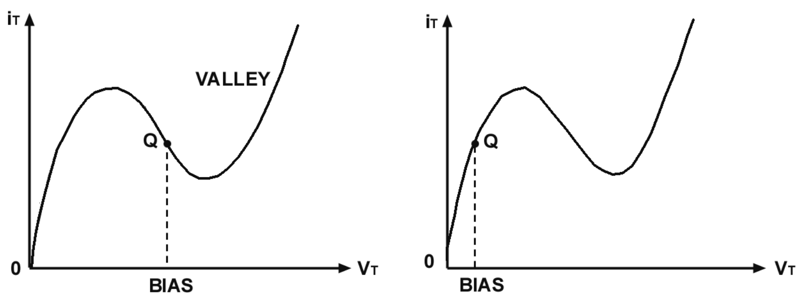



*Fig. 23.22*




(a) (b)

*Fig. 23.23*

**(f) Biasing the Diode**

The tunnel diode has to be biased from some dc source for fixing its Q-point on its characteristic when used as an amplifier or as an oscillator and modulator.

The diode is usually biased in the negative resistance region [Fig. 23.23 (a)]. In mixer and relaxation oscillator applications, it is biased in the positive resistance region near to zero [Fig. 23.23 (b)].

**(g) Applications**

Tunnel diode is commonly used for the following purposes :

1. As an ultrahigh-speed switch due to tunneling mechanism which essentially takes place at the speed of light. It has a switching time of the order of nanoseconds or even picoseconds.
2. As logic memory storage device - due to triple-valued feature of its curve for current.
3. As microwave oscillator at a frequency of about 10 GHz-due to its extremely small capacitance and inductance and negative resistance.
4. In relaxation oscillator circuits - due to its negative resistance.
   In this respect, it is very similar to the unijunction transistor
5. As an amplifier - due to its negative resistance.

**(h) Advantages and Disadvantages**

The advantages of a tunnel diode are :

1. Low noise,
2. Ease of operation,
3. High speed,
4. Low power,
5. Insensitivity to nuclear radiations.

The disadvantages are :

1. The voltage range over which it can be operated properly is 1 V or less,
2. Being a two-terminal device, it provides no isolation between the input and output circuits.

**TUNNEL DIODE OSCILLATOR**

The basic job of an oscillator is to convert dc power into ac power. Ordinarily, we do not expect an ac signal from a circuit which has no input ac source. But the circuit shown in Fig. 23.24 (a) does exactly that as explained below.

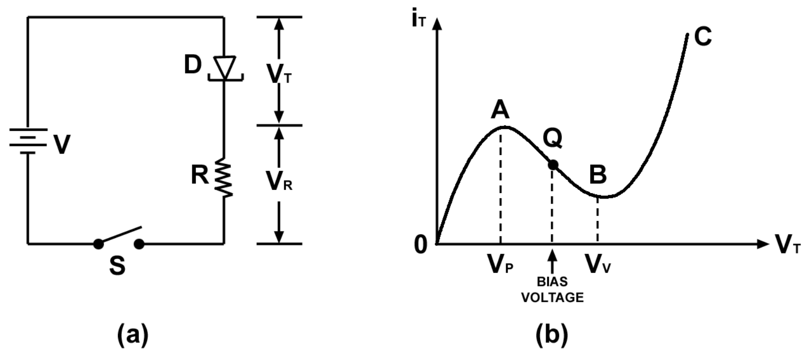



*Fig. 23.24*

The value of R is so selected as to bias the diode D in the negative-resistance region A-B. The working or quiescent point Q is almost at the centre of a curve A-B. When S is closed , the current immediately rises to a value determined by R and diode resistance which are in series. The applied voltage V divides across D and R according to the ratio of their resistances.

However, as $V_T$ exceeds $V_P$, diode is driven into the negative area and its resistance starts to increase. Hence, $V_T$ increases further till it becomes equal to valley voltage $V_V$ (point B). At this point, further increase in $V_T$ drives the diode into the positive-resistance region BC [Fig. 23.24 (b)]. The resulting increase in current now increases $V_R$ but correspondingly decreases $V_T$.

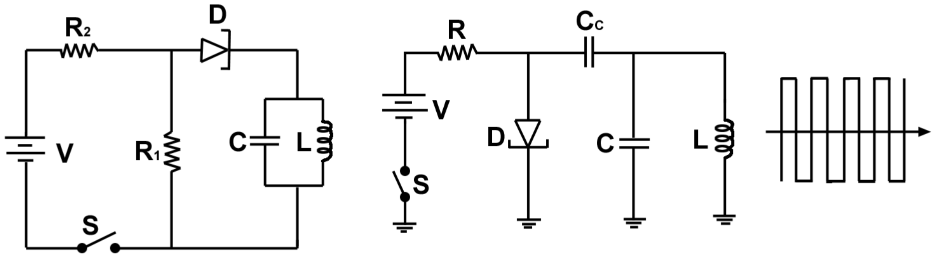



*Fig. 23.25*

It describes one cycle of operation. In this way, the circuit will continue to oscillate back and forth through the negative-resistance region i.e., between points A and B on its characteristic. Its output across R is like a sine wave.

Fig. 23.25 shows a practical circuit drawn in two slightly different ways. Here, $R_2$ sets the proper bias level for the diode whereas $R_1$ (in parallel with the LC tank circuit) sets proper current level for it. The capacitor $C_C$ is the coupling capacitor. As the switch S is closed, the diode is set into oscillations whose frequency equals the resonant frequency of the tank.

**VARACTOR DIODE**

The varactor diode is a semiconductor, voltage-dependent variable capacitor alternatively known as varicap or voltacap or voltage-variable capacitor (VVC) diode. Basically, it is just a reverse-biased junction diode whose mode of operation depends on its transition capacitance ( $C_T$ ). As explained earlier, reverse-biased junctions behave like capacitors whose capacitance $\propto 1/V_R^n$ where n varies from 1/3 to 1/2. As reverse voltage $V_R$ is increased, depletion layer widens thereby decreasing the junction capacitance. Hence, we can change diode capacitance by simply changing $V_R$ . Silicon diodes which are optimised for this variable capacitance effect are called varactors.

The schematic symbol and a simple equivalent circuit for a varactor are shown in Fig. 23.26.
Varactors may be of two types as shown in Fig. 23.27.

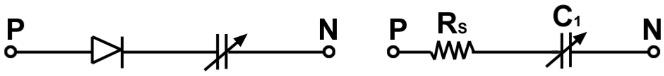



*Fig. 23.26*

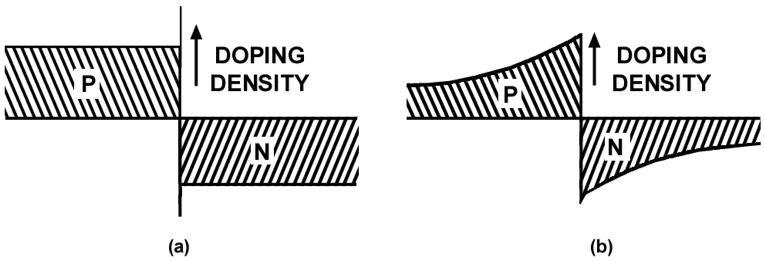



*Fig. 23.27*

The doping profile of the abrupt-junction diode is shown in Fig. 23.27 (a) and that of the hyper abrupt junction diode in Fig. 23.27 (b). The abrupt-junction diode has uniform doping and a capacitive tuning ratio (TR) of 4 : 1. For example, if its maximum transition capacitance is 100 pF and minimum 25 pF, then its TR is 4 : 1 which is not enough to tune a broadcast receiver over its entire frequency range of 550 to 1650 KHz.

The hyper abrupt-junction diode has highest impurity concentration near the junction. It results in narrower depletion layer and larger capacitance. Also, changes in $V_R$ produce larger capacitance changes. Such a diode has a tuning range of 10 : 1, enough to tune a broadcast receiver through its frequency range of nearly 3 : 1.

**Application**

Since the junction capacitance of a varactor is in the pF range, it is suitable for use in high-frequency circuits. Its main applications are as

1. Automatic frequency control device,
2. FM modulator,
3. Adjustable band-pass filter,
4. Parametric amplifier.

**PIN DIODE**

**(a) construction**

It is composed of three sections. There are the usual P-and N-regions but sandwiched between them is an intrinsic layer or I-layer of pure silicon (Fig. 23.28). Being intrinsic (or undoped) layer, it offers relatively high resistance. This high-resistance region gives it two advantages as compared to an ordinary P-N diode. The advantages are :

1. Decrease in capacitance $C_{pn}$ because capacitance is inversely proportional to the separation of P-and N-regions.
   It allows the diode a faster response time. Hence, PIN diodes are used at high frequencies (more than 300 MHz).

2. Possibility of greater electric field between the P- and N-junctions.
   It enhances the electron-hole pair generation thereby enabling PIN diode to process ever very weak input signals.

**(b) Diode Resistance**

1. When forward-biased, it offers a variable resistance $r_{ac} \cong 50 / I$ where I is the dc current in mA . For large dc currents, it would look like a short.
2. When reverse-biased, it looks like an 'open' i.e., it offers infinite resistance in the reverse direction.

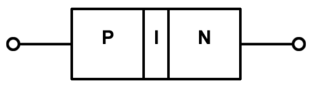



*Fig. 23.28*

**(c) Operation**

***(i) High Frequency Switching***

Its use in electronic high frequency switching is illustrated in Fig. 23.29.

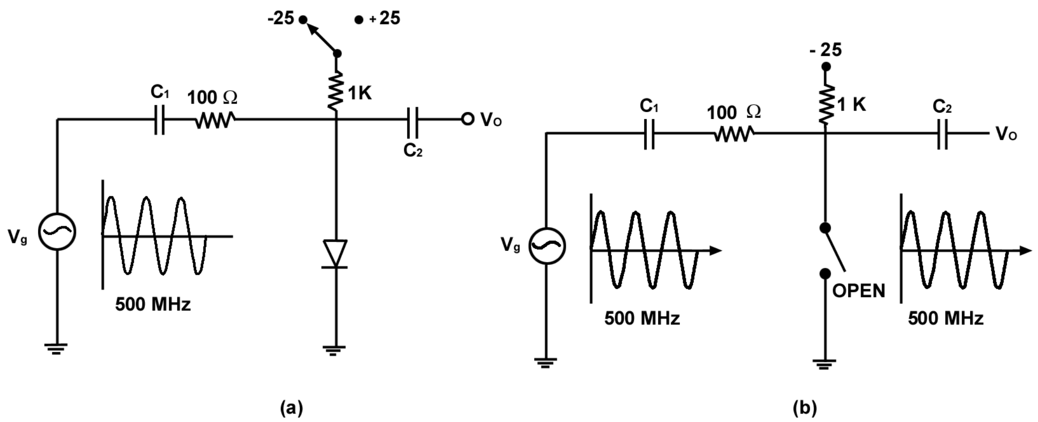



*Fig. 23.29*

When the diode is reverse-biased, it looks like an 'open' as shown in Fig. 23.29 (b). The 500 MHz input signal voltage divides across the series-connected 100 $\Omega$ resistor and the diode in proportion to their resistances. Since the diode has infinite resistance (being open), the entire input signal appears across it. Hence, the whole input signal passes out via coupling capacitance $C_2$ without any attenuation (or loss). When the diode is forward-biased by the +25 V dc source, $I = 25/1K = 25\,mA$ . Hence, diode resistance $r_{ac} = 50/25 = 2\Omega$ as shown by its equivalent circuit in Fig. 23.30. Now, almost all the input signal voltage drops across 100 $\Omega$ resistor and practically none across the 2 $\Omega$ resistance. Hence, there is hardly any signal output.

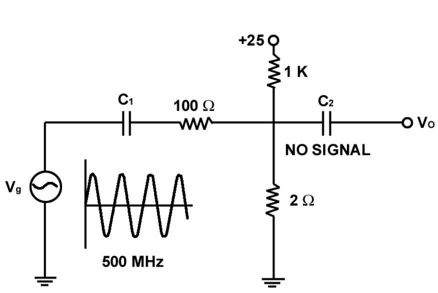



*Fig. 23.30*

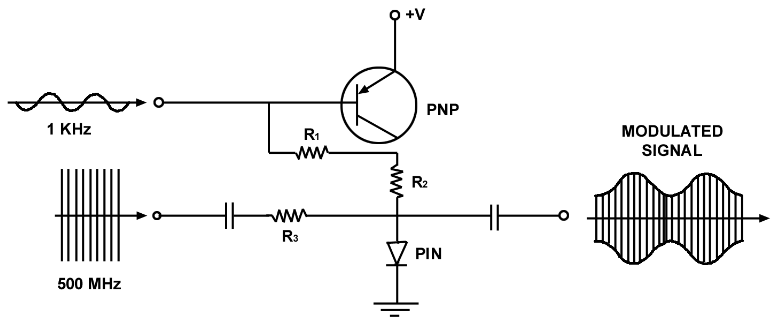



*Fig. 23.31*

In practice, instead of mechanically switching the diode-biasing supply from -25 V to +25 V, a transistor is used to do this switching operation. In this way, we can turn a very high frequency signal (MHz range) OFF and ON with the speed of a transistor switching circuit.

***(ii) Use as AM Modulator***

The way in which the 500 MHz signal is modulated at 1 KHz rate is illustrated in Fig. 23.31. A 1 KHz signal is fed into a PNP transistor where it varies its dc output current at the same rate. This varying dc current is applied as biasing current to the PIN diode as shown in Fig. 23.31. It varies the diode ac resistance as seen by the 500 MHz signal. Hence, the signal is modulated at 1 KHz rate as shown.

**(d) Application**

1. As a switching diode for signal frequencies up to GHz range,
2. As an AM modulator of very high frequency signals.

**SCHOTTKY DIODE**

It is also called Schottky barrier diode and hot-carrier diode. It is mainly used as rectifier at signal frequencies exceeding 300 MHz. It has more uniform junction region and is more rugged than PIN diode.

**(a) Construction**

It is a metal-semiconductor junction diode with no depletion layer. It uses a metal (like gold, silver, platinum, tungsten etc.) on one side of the junction and usually an N-type doped silicon semi-conductor on the other side. The diode and its schematic symbol are shown in fig. 23.32.

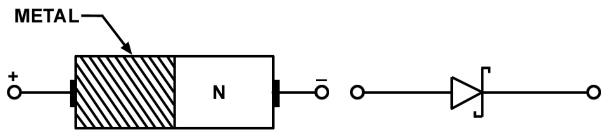



*Fig. 23.32*

**(b) Operation**

When the diode is unbiased, electrons on the N-side have lower energy levels than electrons in the metal. Hence, they cannot surmount the junction barrier (called Schottky barrier) for going over to the metals.

When the diode is forward-biased, conduction electrons on the N-side gain enough energy to cross the junction and enter the metal. Since these electrons plunge into the metal with very large energy, they are commonly called 'hot carriers'. That is why this diode is often referred to as "hot carrier diode".

**(c) Application**

This diode possesses two unique features as compared to an ordinary P-N junction diode :

1. It is a unipolar device because it has electrons as majority carriers on both sides of the junction. An ordinary P-N junction diode is a bipolar device because it has both electrons and holes as majority carriers.
2. Since no holes are available in metal, there is no depletion layer or stored charges to worry about. Hence, Schottky diode can switch OFF faster than a bipolar diode.

Because of these qualities, Schottky diode can easily rectify signals of frequencies exceeding 300 MHz. As shown in Fig. 23.33, it can produce an almost perfect half-wave rectified output.

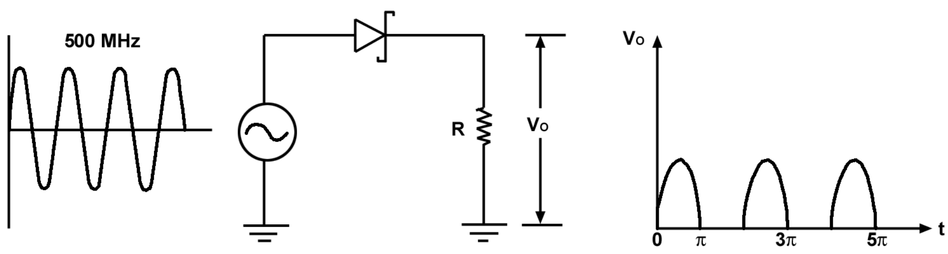



*Fig. 23.33*

The present maximum current rating of the device is about 100 A. It is commonly-used in switching power supplies that operate at frequencies of 20 GHz. Another big advantage of this diode is its low noise figure which is extremely important in communication receivers and radar units etc.

It is also used in clipping and clamping circuits, computer gating, mixing and detecting networks used in communication systems.

**STEP RECOVERY DIODE**

It is another type of Voltage-Variable Capacitor diode having a graded doping profile where doping density decreases near the junction as shown in Fig. 23.34. This results in the production of strong electric fields on both sides of the junction

**(a) Theory**

At low frequencies, an ordinary diode acts as a rectifier. It conducts in the forward direction but not in the reverse direction i.e., it recovers immediately from ON state to the OFF state. However, it is found that when driven forward-to-reverse by a high-frequency signal (above a few MHz), the diode does not recover immediately. Even during the negative half-cycle of the input signal when the diode is reversed-biased, it keeps conducting for a while after which the reverse current ceases abruptly in one step. This reverse conduction is due to the fact that charges stored in the depletion layer during the period of forward bias take time to drain away from the junction.

Fig. 23.35 (a) shows a step-recovery diode being driven by a 20 MHz signal source. As seen from Fig. 23.35 (b), it conducts in the forward direction like any diode. During the reverse half-cycle, we get reverse current due to the draining of the stored charge after which current suddenly drops to zero. It looks as though the diode has suddenly snapped open during the early part of the reverse cycle. That is why it is sometimes called a snap diode.

The step or sudden recovery from reverse current ON to reverse current OFF gives the diode its name.

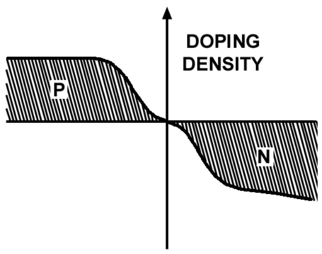



*Fig. 23.34*

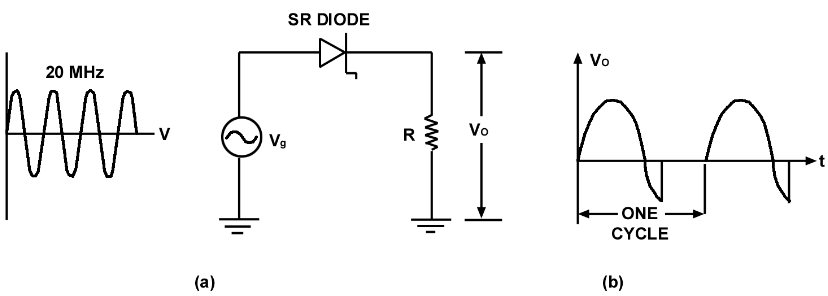



*Fig. 23.35*

**(b) Applications**

Its mains use is in high-frequency harmonic generator circuits as a frequency multiplier as explained below.

It is found that whenever a waveform has sudden step or transition, it contains all the harmonics of the input signal (i.e., multiples of its fundamental frequency). For example, the output waveform of Fig. 23.36 (b) contains waves of frequencies 40 MHz, 60 MHz, 80 MHz and so on.

Fig. 23.36 shows how the output of a step recovery diode can be used to drive a tuning circuit which can be made to tune out all harmonics except one i.e., fifth in this case (100 MHz). With an input signal of 20 MHz, the step recovery diode generates harmonics of different multiple frequencies listed above. However, the resonant L-C circuit is tuned to 5th harmonic of f = 100 MHz. Hence, all except this harmonic are filtered out of the circuit. The signal appearing across R is almost a pure sine wave with f = 100 MHz as shown separately in Fig. 23.36 (b).

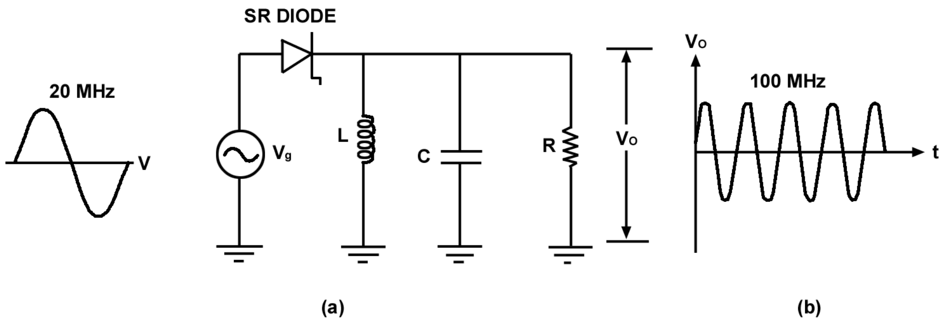



*Fig. 23.36*

Step-recovery diodes are also used in pulse and digital circuits for generating very fast pulses with rise time of less than 1 nanosecond.

■■■

# CHAPTER : 24
# TRIODES

## TRIODE-PHYSICAL CHARACTERISTICS

A typical cut- away sketch of a high-vacuum triode has been shown in Fig. 24.1 As its name implies, a triode contains three active electrodes. The cathode is located at the centre of the tube and is surrounded by a grid which is, in turn, surrounded by an anode. Its symbol is shown in Fig. 24.2 and it is the same as for a diode except for the addition of a dotted line between the anode and the cathode representing the grid. The grid may be in the form of a helix, a squirrel cage or it has apertures or perforations through which electrons can freely pass from cathode to anode. The shape of the anode is influenced by the shape of cathode and grid and is usually round, elliptical or even rectangular.

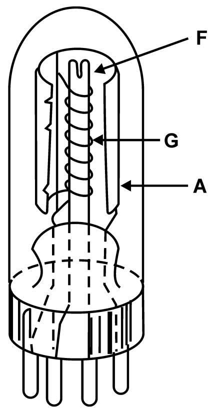



*Fig. 24.1*

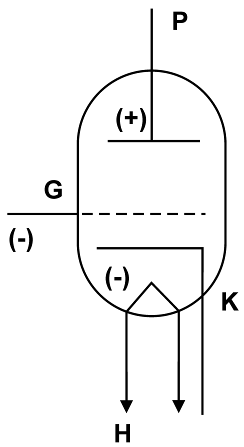



*Fig. 24.2*

The primary function of the additional third electrode i.e. grid is to serve as an electrostatic screen and thus partially shield the cathode from the electrostatic field of the anode. The operation of the grid can be explained thus :

First, let it be pointed out that grid is directly situated in the path of the electronic beam and secondly, it is nearer to cathode advantageous position. When the grid is given some negative potential with respect to the cathode, it repels back or retards the flow of electrons towards the plate, thereby reducing the plate current. Because of its advantageous position as pointed out, even with a small negative potential, it is capable of exerting greater controlling effect on plate current than anode itself even with a much larger positive potential. Because of this current-controlling property of the gird, it is often referred to as control grid. As the negative potential of the grid (called grid bias) is increased, $I_p$ (plate current) decreases continuously till ultimately it is reduced to zero. Grid-bias corresponding to zero plate current is known as cutoff potential of the grid. Its value can be found by putting $I_p = 0$ in the triode equation given below. If on the other hand, it is made positive relative to cathode, it helps the anode in pulling the electrons up by partially neutralizing the negative space charge. Hence, $I_p$ is increased. It is found that small changes in grid voltage produce such changes in plate current which can be produced by comparatively large changes in plate voltage. It is because of this that a triode possesses that valuable property of amplification. If a change of 2 volts on the grid produces the same change in plate current as produced by a change of 50 volts in plate voltage, then the amplification factor ($\mu$) is 50/2 = 25. It means that the grid (because of its advantageous positioning) is 25 times more effective than anode in controlling the plate current.

From the above discussion it is obvious that the plate current depends not only on cathode temperature and plate voltage $V_p$ but on grid voltage $V_q$ also. In the region where plate current is principally limited by space-charge, the expression for plate current is given by

$$I_p = K ( V_q + V_p / \mu )^{3/2}$$

where K is a constant depending upon the geometry of the tube and h is the amplification factor. The above equation can also be put in an alternate form

$$I_p = A ( V_p + \mu V_q )^{3/2}$$

where A is another constant equal to $K/\mu^{3/2}$.

### Electrical Characteristics of a Triode

A triode has static as well as dynamic characteristics. The static characteristics give relation between different parameters of a triode when different dc potential are applied to its electrodes. The dynamic characteristics are the values obtained with ac voltage applied to the control grid when other electrodes have dc potential. Hence this characteristic indicates the performance capabilities of the triode under actual working conditions.

The three static characteristics of a triode are:

1. ***$V_p$ / $I_p$ characteristic.*** It also called plate or anode characteristic.
2. ***$V_g$ / $V_p$ characteristic.*** It is also called transfer or mutual characteristic.
3. ***$V_g$ / $V_p$ characteristic.*** It is also called constant-current characteristic.

**Plate Characteristic of a Triode**

For obtaining this characteristic, dc grid voltage $V_g$ is kept constant at a given dc value whereas plate voltage $V_p$ is changed and corresponding values of $I_p$ (plate current) are noted. The results so obtained are plotted in Fig .24.3 which are typical for a triode.

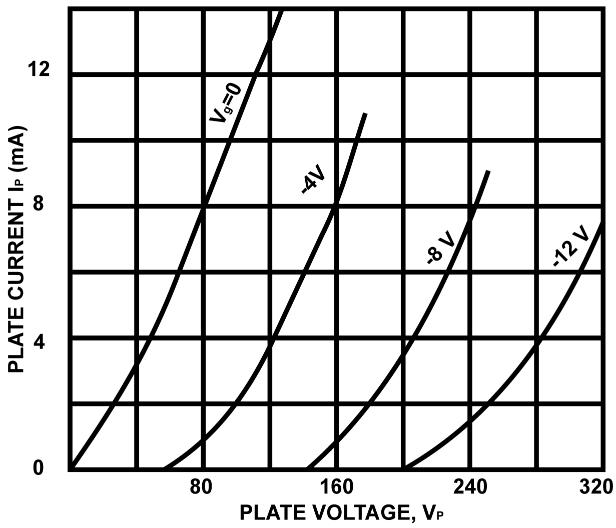



*Fig. 24.3*

**Transfer Characteristic**

This characteristic shows variation of $I_p$ with $V_g$ with fixed values of $V_p$. Different characteristics corresponding to different plate potential are shown in Fig. 24.4

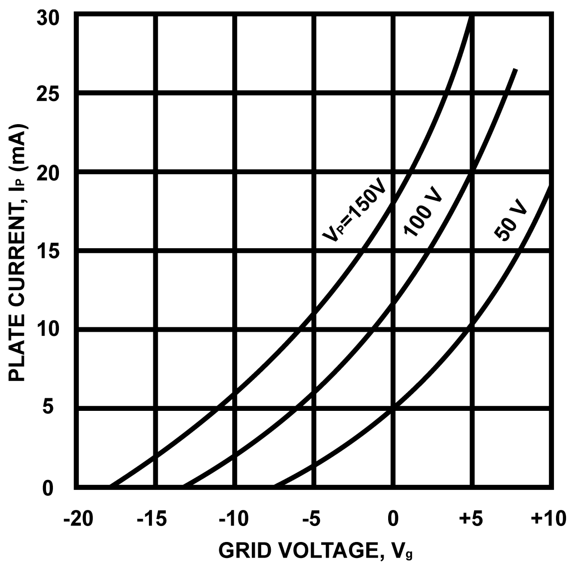



*Fig. 24.4*

**Constant-current Characteristic**

It shows the relation between plate voltage and grid voltage for keeping plate current constant as a given value, Fig.24.5.

**TRIODE COEFFICIENTS**

Three coefficients or parameters of a triode are as under:

1. ***Amplification Factor ( μ ) :-*** It is given by the ratio of the change in plate voltage to the change in grid voltage for keeping the plate current constant.

$$\mu = (\Delta V_p / \Delta V_g) \mid I_p \text{ constant}$$

It is a dimension less number and is represented by the slope of the constant-current characteristic.

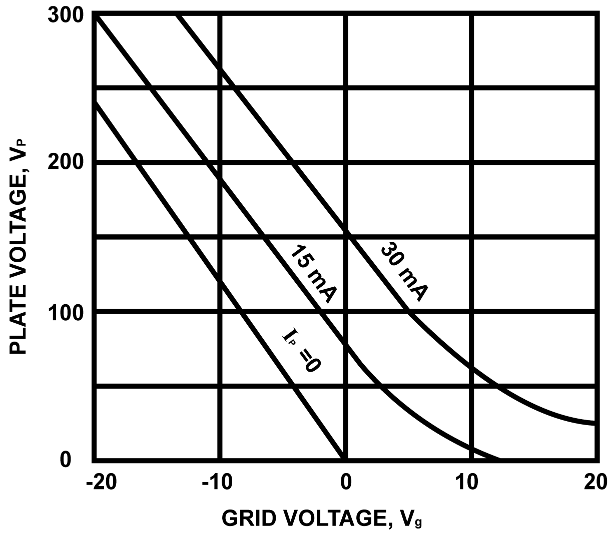



*Fig. 24.5*

2. ***Plate Resistance :-***. It is given by the ratio of the change in plate voltage to the change in plate current for a constant grid voltage.
   $R_p = (\Delta V_p / \Delta I_p) \mid V_g$ constant.
   It is also called ac plate resistance and is equal to the reciprocal of the plate characteristic.

3. ***Mutual Conductance :-*** It is given by the ratio of change in plate current to the change in grid voltage for keeping plate voltage constant at the given value.
   $g_m = (\Delta I_p / \Delta V_g) \mid V_p$ constant
   Its unit is siemens.

***Interrelation of Three Coefficients***

We have seen above that,

$$\mu = (\Delta V_p / \Delta V_g) = (\Delta V_p / \Delta I_p) X (\Delta I_p / \Delta V_g) = R_p X g_m$$

$$\mu = R_p X g_m$$

**TRIODE AS AN AMPLIFIER**

A triode can act as an amplifier because any small change produced in $V_g$ produces a big change in $I_p$ and hence large potential drop across the load resistance in the plate circuit.

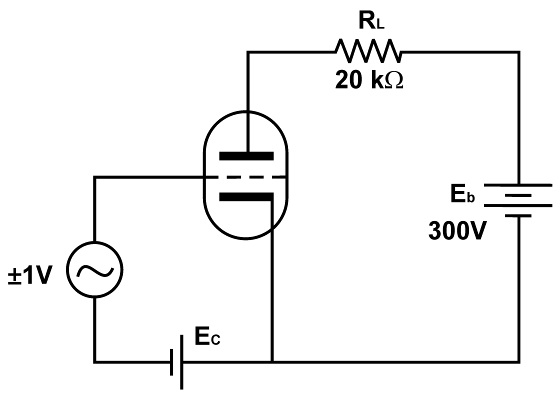



*Fig. 24.6*

A simple triode amplifier circuit is shown in fig. 24.6 The dc plate potential is supplied by the battery $E_b$ and the dc grid bias is supplied by the battery $E_c$. A small ac signal of ± 1 V is applied between the grid and cathode. This signal makes the grid potential either less or more thereby increasing or decreasing $I_p$. This changing $I_p$ increases or decreases the voltage drop across the 20 k Ω plate load resistance. Suppose that a variation of ± 1 V in $V_g$ produces a variation of ± 50 V in $V_p$. In that case, the amplification factor of the triode is = 50/1 = 50.

# CHAPTER : 25
# BIPOLAR JUNCTION TRANSISTOR

**THE BIPOLAR JUNCTION TRANSISTOR**

Basically, it consists of two back-to-back P-N junctions manufactured in a single piece of a semiconductor crystal. These two junctions give rise to three regions called emitter, base and collector. As shown in Fig, 25.1 a junction transistor is simply a sandwitch of one type of semiconductor material between two layers of the other type. Fig . 25.1(a) shows a layer of N-type material sandwiched between two layers of P-type material. It is described as PNP transistor. Fig. 25.1 (b) Shows an NPN transistor consisting of a layer of P-type material sandwiched between two layers of N-type material.

The emitter, base and collector are provided with terminals which are labelled as E,B and C. The two junctions are: emitter - base ( E/B ) junction and collector-base ( C/B ) junction.

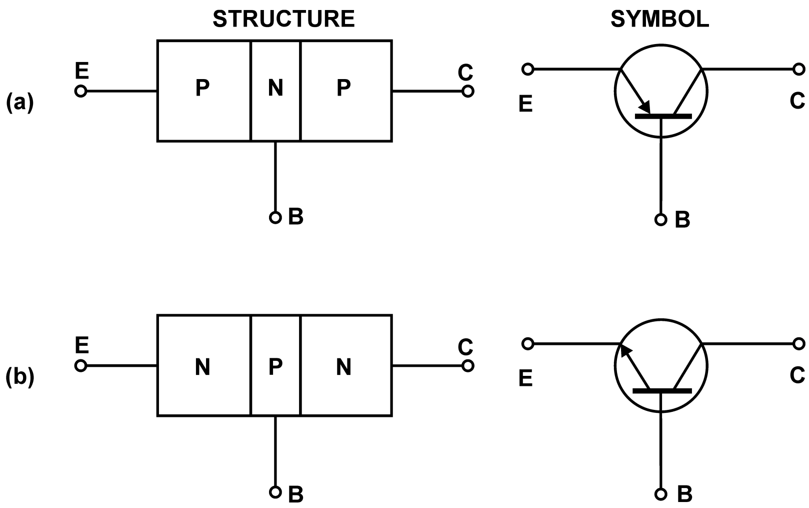



*Fig. 25.1*

The symbols employed for PNP and NPN transistors are also shown in Fig. 25.1. The arrowhead is always at the emitter ( not at the collector) and in each case, its direction indicates the conventional direction of current flow. For a PNP transistor, arrowhead points from emitter to base meaning that emitter is positive with respect to base (and also with respect to collector) . For an NPN transistor, it points from base to emitter meaning that base (and collector as well) is positive with respect to emitter.

**Emitter**

It forms the left-hand section or region of the transistor as shown in Fig. 25.1. It is more heavily doped than any of the other regions because its main function is to supply majority charge carriers (either electrons or holes) to the base.

**Base**

It forms the middle section of the transistor. It is very thin ( $10^{-6}$ m ) as compared to either the emitter or collector and is very lightly doped.

**Collector**

It forms the right-hand side section or region of the transistor as shown in fig.25.1, and its main function is ( as indicated by its name ) to collect majority charge carries coming from the emitter and passing through the base.

In most transistors, collector region is made physically larger than the emitter region because it has to dissipate much greater power.

It may be noted, that transistors are made by growing, alloying or diffusing processes.

**Transistor Biasing**

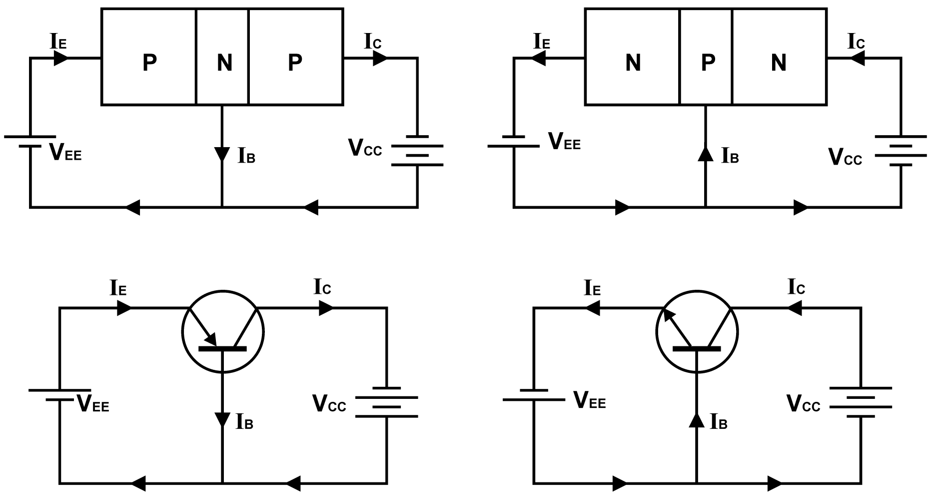
For proper working of a transistor, it is essential to apply voltage of correct polarity across its two junctions. It is worthwhile to remember that for normal operation.

1. Emitter-base junction is always forward-biased and
2. Collector-base junction is always reverse-biased

This type of biasing is known as FR biasing.

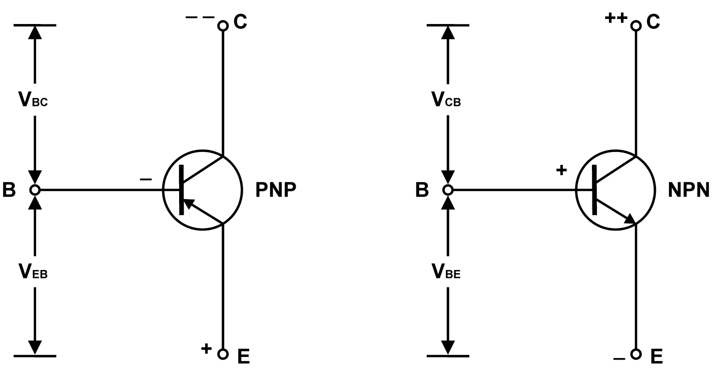
In Fig. 25.2, two batteries respectively provide the dc emitter supply voltage $V_{EE}$ and collector supply voltage Vcc for properly biasing the two junctions of the transistor. In fig 25.2 (a), Positive terminal of $V_{EE}$ is connected to P-type emitter in order to repel or push holes into the base.

The negative terminal of $V_{cc}$ is connected to the collector so that it may attract or pull holes through the base. Similar considerations apply to the NPN transistor of fig. 25.2 (b). It must be remembered that a transistor will never conduct any current if its emitter base junction is not forward biased.




(a) (b)

*Fig. 25.2*




(a) (b)

*Fig. 25.3*

**Important Biasing Rule**

For a PNP transistor, both collector and base are negative with respect to the emitter (the letter N of negative being the same as the middle letter of PNP ). Of course, collector is more negative than base Fig. 25.3 (a). Similarly, for an NPN transistor, both collector and base are positive with respect to the emitter ( the letter P of Positive being the same as the middle letter of NPN). Again, collector is more positive than the base as shown in Fig. 25.3 (b)

**Transistor Currents**

The three primary currents which flow in a properly-biased transistor are $I_E$, $I_B$ and $I_c$. In Fig. 25.4 (a) are shown the directions of flow as well as relative magnitude of these currents for a PNP transistor connected in the common-base mode. It is seen that

$$I_E = I_B + I_c$$

It is seen that a small part (about 1-2%) of emitter current goes to supply base current and the remaining major part (99-98%) goes to supply collector current.

Moreover, $I_E$ flows into the transistor whereas both $I_B$ and $I_c$ flow out of it.

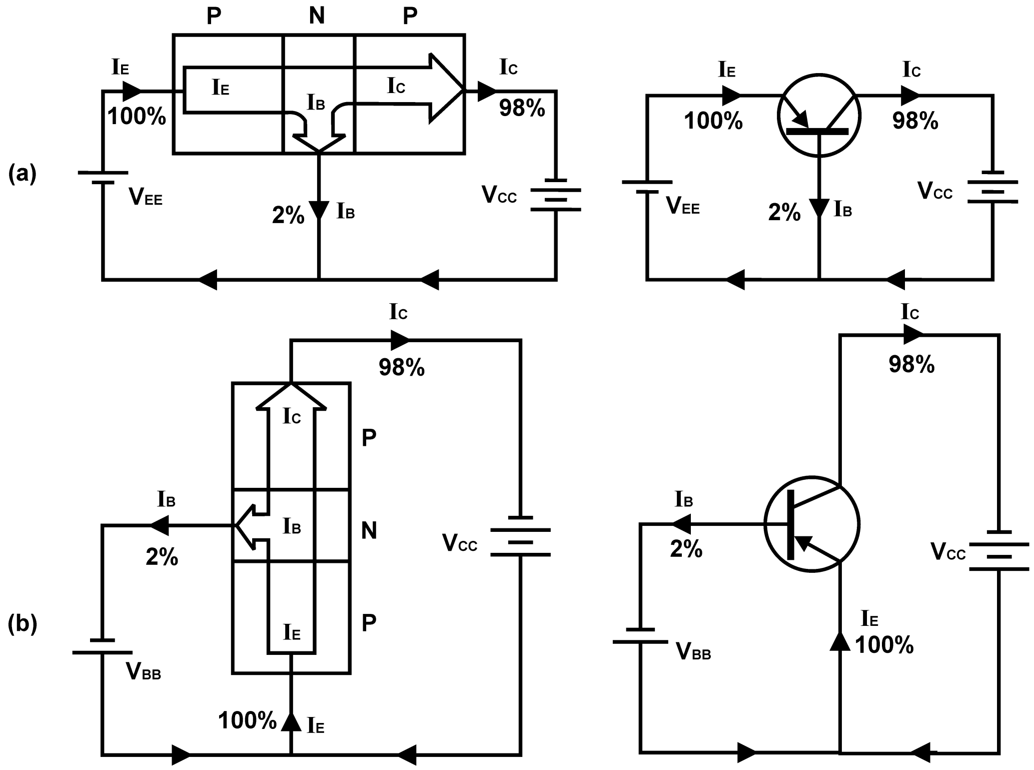



*Fig. 25.4*

Fig.25.4 (b) shows the flow of currents in the same transistor when connected in the common-emitter mode. It is seen that again

$$I_E = I_B + I_c$$

By normal convention, currents flowing into a transistor are taken as positive whereas those flowing out of it are taken as negative. Hence, $I_E$ is positive whereas both $I_B$ and $I_c$ are negative. Applying Kirchhoff's Current Law, We have

$$I_E + (-I_B) + (-I_c) = 0$$

or $$I_E - I_B - I_c = 0$$

$$I_E = I_B + I_c$$

This statement is true regardless of transistor type or transistor configuration.

**Note**. For the time being, we have not taken into account the leakage currents which exist in a transistor.

**Summing Up**

The four basic guide posts about all transistor circuits are:

1. Conventional current flows along the arrow whereas electrons flow against it.
2. E/B junction is always forward-biased
3. C/B junction is always reverse-biased
4. $I_E = I_B + I_c$

**Transistor Circuit Configurations**

Basically, there are three types of circuit connections (called configurations) for operating a transistor

1. Common-base ( CB )
2. Common-emitter (CE)
3. Common-collector ( CC)

The term 'common' is used to denote the electrode that is common to the input and output circuits. Because the common electrode is generally grounded, these modes of operation are frequently referred to as grounded-base, grounded-emitter and grounded-collector configurations as shown in Fig. 25.5, 25.7 and 25.8.

**CB Configuration**

In this configuration Fig. 25.5, emitter current $I_E$ is the input current and collector current $I_c$ is the output current. The input signal is applied between the emitter and base whereas output is taken out from the collector and base.

The ratio of the collector current to the emitter current is called dc alpha ($\alpha_{dc}$) of a transistor.

$$\alpha_{dc}* = \frac{I_C - I_{CBO}}{I_E} \quad \text{or} \quad I_C = -\alpha_{dc}.I_E **$$

If we write $\alpha_{dc}$ simply as $\alpha$, then

$$\alpha = +\frac{I_C}{I_E} \quad \text{or} \quad I_c = \alpha\, I_E$$

It is also called forward current transfer ratio ($-h_{fb}$). In $h_{fb}$, subscript 'f' stands for forward and 'b' for common base. The negative sign merely indicates that emitter and collector currents flow in opposite directions. The subscript dc on $\alpha$ signifies that this ratio is defined from dc values of $I_c$ and $I_E$.

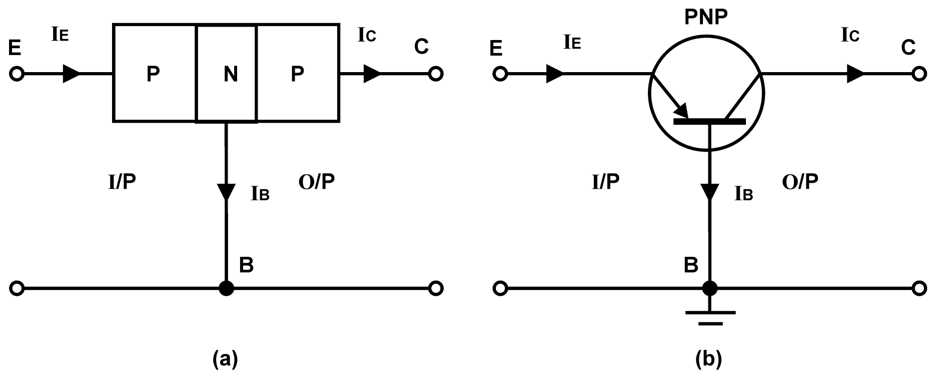



*Fig. 25.5*

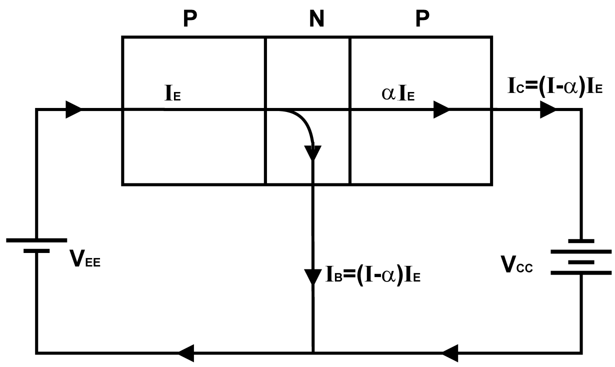



*Fig. 25.6*

**Note :-** * More accurately, $\alpha_{dc} = \frac{I_C - I_{CBO}}{I_E}$

** Negative sign has been omitted since we are here concerned with only magnitude of the currents involved.

The $\alpha$ of a transistor is a measure of the quality of a transistor : higher the value of $\alpha$, better the transistor in the sense that collector current more closely equals the emitter current. Its value ranges from 0.95 to 0.999. Obviously, it applies only to CB configuration of a transistor. As seen from above.

$$I_c = \alpha\, I_E$$

Now,
$$I_B = I_E - I_c$$
$$= I_E - \alpha\, I_E$$
$$I_B = (1-\alpha) I_E$$

Incidentally, there is also an ac $\alpha$ for a transistor. It refers to change in collector current to change in emitter current.

$$\alpha_{ac} = \frac{\Delta I_C}{\Delta I_E}$$

It is also known as short-circuit gain of a transistor and is written as $h_{fb}$. It may be noted that upper case subscripts FB indicate dc value whereas lower case subscripts ' fb ' indicate ac value. For all practical purposes, $\alpha_{dc} = \alpha_{ac}$

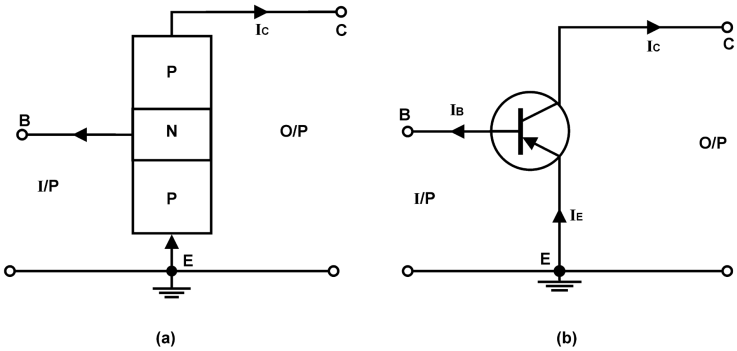



*Fig. 25.7*

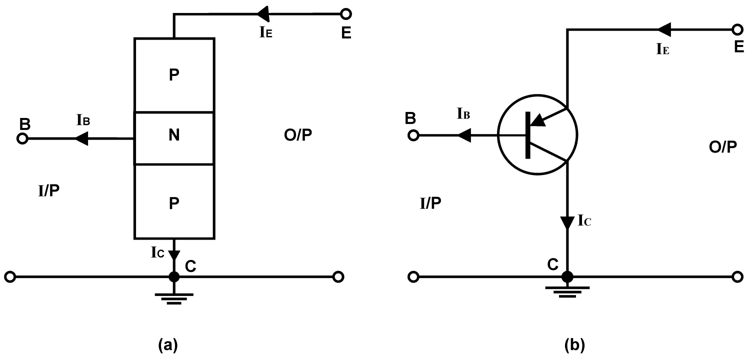



*Fig. 25.8*

**CE Configuration**

Here, input signal is applied between the base and emitter and output signal is taken out from collector and emitter circuits. As seen form Fig . 25.7 (b), $I_B$ is the input current and $I_c$ is the output current.

The ratio of dc collector current to dc base current is called dc beta ($\beta_{dc}$) or just $\beta$ of the transistor.

$$\beta = \frac{I_C}{I_B} \quad \text{or} \quad I_C = \beta\, I_B$$

It is also called common-emitter dc forward transfer ratio and is written as $h_{fe}$. It is possible for $\beta$ to have as high a value as 500.

While analysing ac operation of a transistor, we use ac $\beta$ which is given by

$$\beta_{ac} = \frac{\Delta I_C}{\Delta I_B}$$

It is also written as $h_{fe}$. Since Ic and $I_B$ is positive unlike $\alpha$ which is negative.

**Relation Between $\alpha$ and $\beta$**

$$\beta = \frac{I_C}{I_B} \qquad \text{and} \qquad \alpha = \frac{I_C}{I_E}$$

$$\therefore \quad \frac{\beta}{\alpha} = \frac{I_E}{I_B}$$

Now, $I_B = I_E - I_c$

$\beta = [(I_c)/(I_E - I_c)]$

$= [(I_c/I_E)/\{(I_E/I_E) - (Ic/I_E\}]$

$\beta = \alpha/(1-\alpha)$ (i)

Cross-multiplying the above equation and simplifying it, we get

$\beta(1-\alpha) = \alpha$ or $\beta = \alpha(1+\beta)$

$\alpha = \beta/(1+\beta)$ (ii)

It is seen from the above two equations that

$1-\alpha = 1/(1+\beta)$

**CC Configuration**

In this case ( Fig. 25.8), input signal is applied between base and collector and output signal is taken out from emitter-collector circuit. Obviously, conventionally speaking, here $I_B$ is the input current and $I_E$ is the output current as shown in Fig. 25.8 The current gain of the circuit is

$(I_E/I_B) = (I_E/I_c).(I_c/I_B) = (1/\alpha).\ \beta$

$(\beta/\alpha) = \beta/\{\beta/(1+\beta)\} = (1+\beta)$

It means that output current is $(1+\beta)$ times the input current.

**Relations Between Transistor Currents**

While deriving various equations, following definitions should be kept in mind.

$\alpha = Ic/I_E$ and $\beta = Ic/I_B$

Also, $\alpha = \beta/(1+\beta)$ and $\beta = \alpha/(1-\alpha)$

1. $Ic = \beta I_B = \alpha I_E = \beta/(1+\beta).I_E$
2. $I_B = (Ic/\beta) = I_E/(1+\beta) = (1-\alpha).I_E$
3. $I_E = (Ic/\alpha) = [(1+\beta)/\beta]\, Ic$
   $= (1+\beta)I_B = I_B/(1-\alpha)$
4. The three transistor dc currents always bear the following ratio :
   $I_E : I_B : I_c :: 1 : (1-\alpha) : \alpha$
   Incidentally, it may be noted that for ac currents, small letters $i_e$, $i_b$ and $i_c$ are used.

**LEAKAGE CURRENTS IN A TRANSISTOR**

**CB Circuit**

Consider the CB transistor circuit shown in Fig. 25.9 The emitter current (due to majority carriers) initiated by the forward biased emitter-base junction is split into two parts :-

i) $(1-\alpha)\ I_E$ which becomes base current $I_B$ in the external circuit and

ii) $\alpha\ I_E$ which becomes collector current $I_c$ in the external circuit.

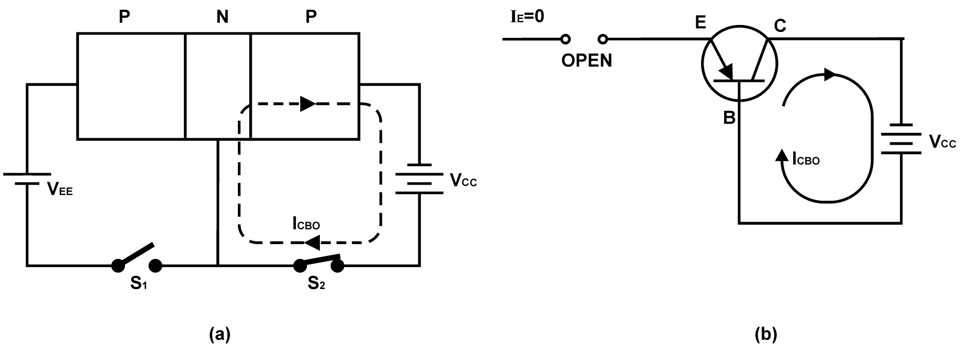



*Fig. 25.9*

As mentioned earlier through C/B junction is reverse-biased for majority charge carries (i.e. holes in this case), it is forward-biased so far as thermally-generated minority charge carries (i.e. electrons) are concerned. This current flows even when emitter is disconnected from its dc supply as shown in Fig 25.9 (a) where switch $S_1$ is open. It flows in the same direction as the collector current of majority carriers. It is called leakage current $I_{CBO}$. The subscripts CBO stand for ' current from Collector to Base with emitter Open. Very often, it is simply written as $I_{co}$.

It should be noted that

i) $I_{CBO}$ is exactly like the reverse saturation current $I_S$ or $I_o$ of a reverse-biased diode.

ii) $I_{CBO}$ is extremely temperature-dependent because it is made up of thermally-generated minority carries. As mentioned earlier, $I_{CBO}$ doubles for every 10ºC rise in temperature for Ge and 6ºC for Si.

If we take into account the leakage current, the current distribution in a CB transistor circuit becomes as shown in Fig.25.10 both for PNP and NPN type transistors.

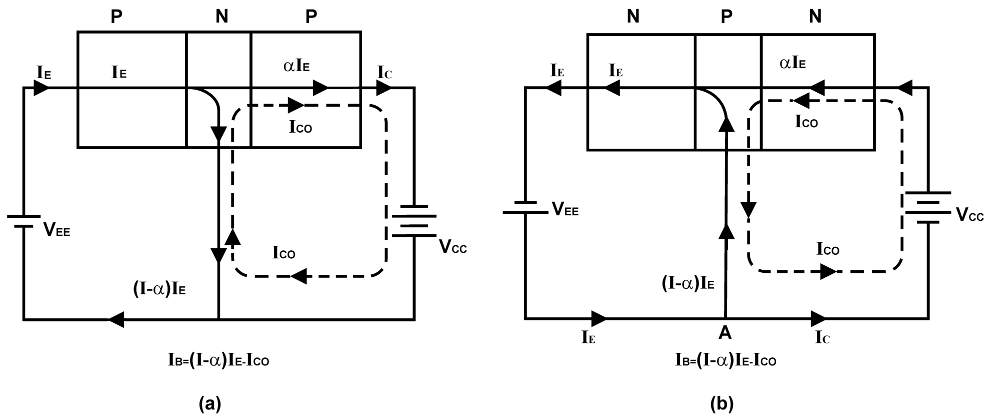



*Fig. 25.10*

It is seen that total collector current is actually the sum of two components

i) current produced by normal transistor action i.e. component controlled by emitter current. Its value is $\alpha I_E$ and is due to majority charge carriers.

ii) temperature-dependent leakage current $I_{co}$ due to minority carriers

$$I_c = \alpha\, I_E + I_{co}$$

$$= \underset{\text{majority}}{\alpha I_E} + \underset{\text{minority}}{I_{co}}$$

It can also be proved that

$$I_c = (\, I_{co} / 1\text{-}\alpha\,) + (\alpha I_B / 1-\alpha\,)$$

*Note:-* *Actually, electrons (which form minority charge carriers in collector) flow from negative terminal of collector battery, to collector, then to base through C/B junction and finally to positive terminal of $V_{cc}$. However, conventional current flows in the opposite direction as shown by the dotted line in fig. 25.10 (a)

In view of the above, it would be appreciated that

$$\alpha = -[(-\,Ic + Ico) / (I_E)] \cong -\,Ic / I_E.$$

It can also be proved that

$$I_B = (I - \alpha\,)I_E - I_{co}$$

**CE Circuit**

In Fig. 25.11 is shown a grounded-emitter circuit of an NPN transistor whose base lead is open. It is found that despite $I_B = O$, there is a leakage current from collector to emitter . It is called $I_{CEO}$, the subscripts CEO standing for ' Collector to Emitter with base Open.

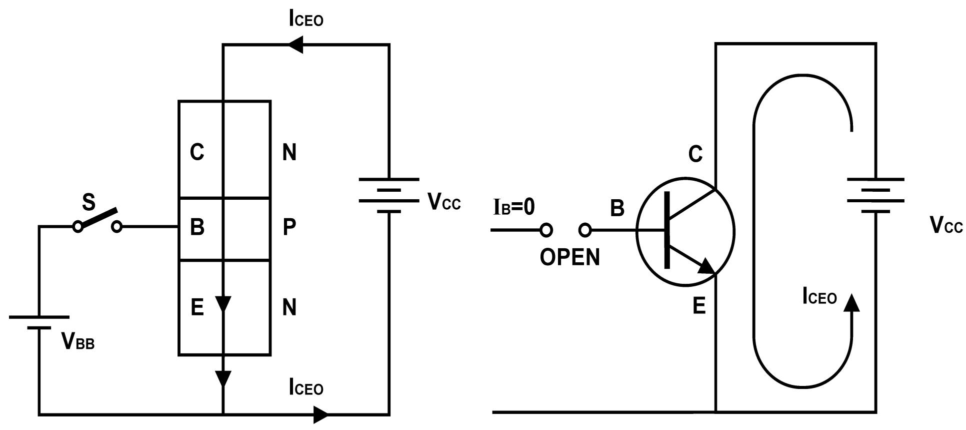



*Fig. 25.11*

Taking this leakage current into account, the current distribution through a CE circuit becomes as shown in Fig. 25.12

$$I_C = \beta I_B + I_{CEO} = \beta I_B + (1+\beta)\, I_{co}$$

$$= \beta I_B + I_{co}/(1\text{-}\alpha\,) \qquad = \alpha I_E + (1-\alpha\,)\, I_{CEO}$$

$$= [(\alpha I_B\,) / (1-\alpha)] + [(I)_{co}/(\,1\text{-}\alpha)]$$

Now, $\beta I_B = \alpha I_E$. Substituting this value above equation we get,

$I_c = \alpha I_E + I_{CEO}$

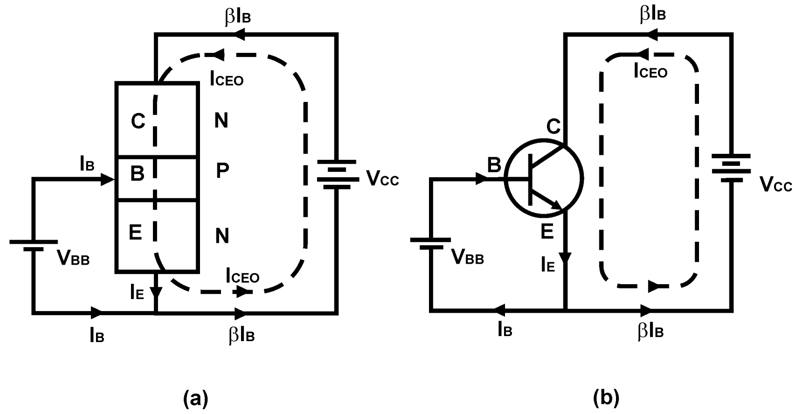



*Fig. 25.12*

Also, $I_B = I_E - I_c$

Substituting the value of $I_c$ from above, we have

$I_B = I_E - \alpha I_E - I_{CEO} = (1-\alpha) I_E - I_{CEO}$

**TRANSISTOR STATIC CHARACTERISTICS**

These are the curves which represent relationships between different dc currents and voltage of a transistor. These are helpful in studying the operation of a transistor when connected in a circuit. The three important characteristics of a transistor are.

1. Input characteristic
2. Output characteristic
3. Constant current transfer characteristic.

**COMMON BASE TEST CIRCUIT**

The static characteristics of an NPN transistor connected in common-base configuration can be determined by the use of the test circuit shown in Fig. 25.13 milliameters are included in series with the emitter and collector circuits for measuring $I_E$ and $I_c$. Similarly, voltmeters are connected across E and B to measure voltage $V_{BE}$* and across C and B to measure $V_{CB}$*. The two potentiometer resistors $R_1$ and $R_2$ supply variable voltage from the collector and emitter dc supplies.

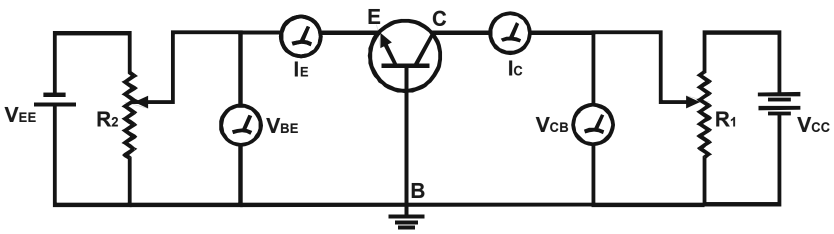



*Fig. 25.13*

**Common Base Static Characteristics**

***(a) Input characteristic***

.It shows how $I_E$ varies with $V_{BE}$ when voltage $V_{CB}$ is held constant. The method of determining the characteristic is as follows:

First, voltage $V_{CB}$ is adjusted to a suitable value with the help of $R_1$. Next, voltage $V_{BE}$ is increased in a number of discrete steps and corresponding values of $I_E$ are noted from the milliammeter connected for the purpose. When plotted, we get the input characteristics shown in Fig 25.14., one for Ge and the other for Si. Both curves are exactly similar to the forward characteristic of a P-N diode which , in essence, is what the emitter-base junction is.

This characteristic may be used to find the input resistance of the transistor. Its value is given by the reciprocal of its slope.

$R_{in} = (\Delta V_{BE} / \Delta I_E) \ / \ V_{CB}$ constant

Since the characteristic is initially non-linear, $R_{in}$ will vary with the point of measurement. Its value over linear part of the characteristic is about 50 Ω but for low values of $V_{BE}$ it is considerably greater. The change in $R_{in}$ with change in $V_{BE}$ gives rise to distortion of signals handled by the transistor.

**Note:** *In these subscripts, first letter indicates that electrode which is at higher potential as compared to the other electrode.

This characteristic is hardly affected by changes either in $V_{CB}$ or temperature.

Note: Strictly speaking, both $I_E$ and $V_{BE}$ should be shown negative for an NPN transistor.

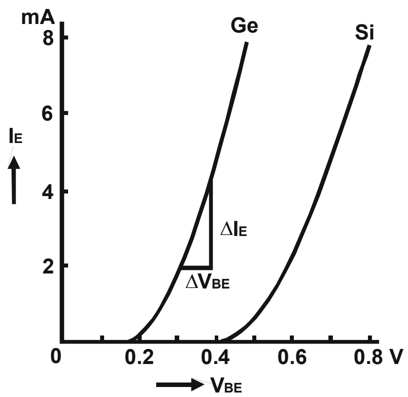



*Fig. 25.14*

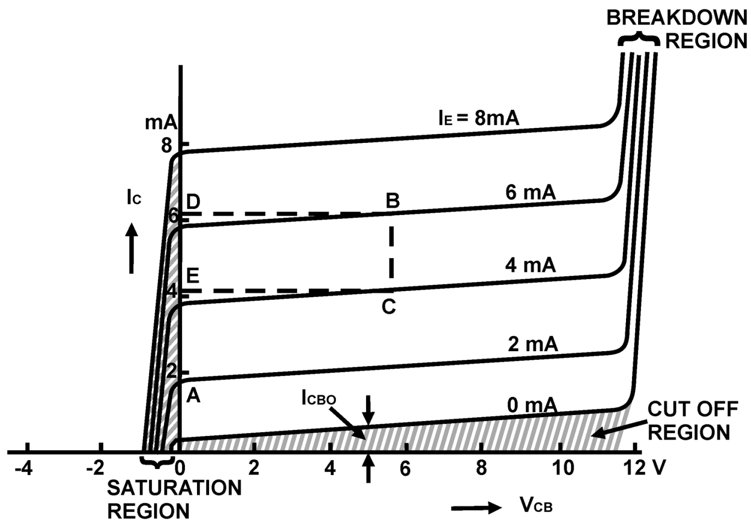



*Fig. 25.15*

### *(b) Output Characteristic*

It shows the way $I_c$ varies with $V_{CB}$ when $I_E$ is held constant.
The method of obtaining the characteristic is as follows :

First, movable contact on $R_2$ is changed to get a suitable value of $V_{BE}$ and hence $I_E$. While keeping $I_E$ constant at this value, $V_{CB}$ is increased from zero in a number of steps and the corresponding collector current $I_c$ that flows is noted.

Next, $V_{CB}$ is reduced back to zero, $I_E$ is increased to a value a little higher than before and the whole procedure is repeated. In this way, whole family of curves is obtained, a typical family being shown in Fig. 25.15

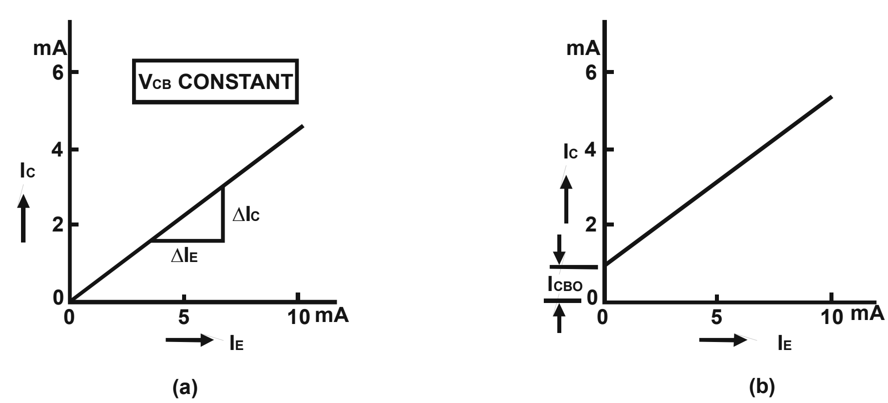
1. The reciprocal of the slope of the near horizontal part of the characteristic gives the output resistance $R_{out}$ of the transistor which it would offer to an input signal . Since the characteristic is linear over most of its length (meaning that $I_c$ is virtually independent of $V_{CB}$), $R_{out}$ is very high, a typical value being 500 kΩ.

   $R_{out} = 1 / (\Delta I_c / \Delta V_{CB}) = (\Delta V_{CB} / \Delta I_c)$

2. It is seen that $I_c$ flows even when $V_{CB} = 0$. For example, it has a value = OA corresponding to $I_E = 2$ mA as shown in fig. 15. It is due to the fact that electrons are being injected into the base under the action of forward-biased E/B junction and are being collected by the collector due to the action of the internal junction voltage at the C/B junction . For reducing $I_c$ to zero, it is essential to neutralize this potential barrier by applying a small forward bias across C/B junction.
3. Another important feature of the characteristic is that a small amount of collector current flows even when emitter current $I_E = 0$. As we know it is collector leakage current $I_{CBO}$.
4. This characteristic may be used to find $\alpha_{ac}$ of the transistor as shown in Fig. 25.15.

   $\alpha_{ac} = (\Delta Ic / \Delta I_E) = (DE/BC)$
   $= 6.2 - 4.3 / 2 = 0.95$

5. Another point worth noting is that although $I_c$ is practically independent of $V_{CB}$ over the working range of the transistor, yet if $V_{CB}$ is permitted to increase, $I_C$ eventually increases rapidly due to avalanche breakdown as shown.




(a) (b)

*Fig. 25.16*

***Current Transfer Characteristic***

It shows how $I_c$ varies with changes in $I_E$ when $V_{CB}$ is held constant. For drawing this characteristic, first $V_{CB}$ is set to a convenient value and then $I_E$ is increased in steps and corresponding values of $I_c$ noted. A typical transfer characteristic is shown in Fig.25.16 (a). fig.25.16 (b). shows a more detailed view of the portion near the origin.

$$\alpha_{ac} = \Delta I_C / \Delta I_E$$

Usually, $\alpha_{ac}$ is found from output characteristic rather than from this characteristic.

It may be noted in the end that CB connection is rarely employed for audio-frequency circuits because firstly, it current gain is less than unity and secondly, its input and output resistances are so different.

Note: Strictly speaking, for an NPN: transistor, $I_C$ should be positive but $I_E$ should be negative.

**COMMON EMITTER TEST CIRCUIT**

The static characteristics of a transistor connected in CE configuration may be determined by the use of circuit diagram shown in Fig. 25.17 A milliammeter (or a micro ammeter in the case of a low-power transistor) is connected in series with the base to measure $I_B$. Similarly, an ammeter is included in the collector circuit to measure $I_C$. A voltmeter with a typical range of 0-1V is connected across base and emitter terminals for measuring $V_{BE}$.

Potentiometer $R_2$ connected across d.c. supply $V_{BB}$ is used to vary $I_B$ and $V_{BE}$. A second voltmeter with a typical voltage of range 0-20 V is connected across collector-emitter terminals to measure the output collector-emitter voltage $V_{CE}$.

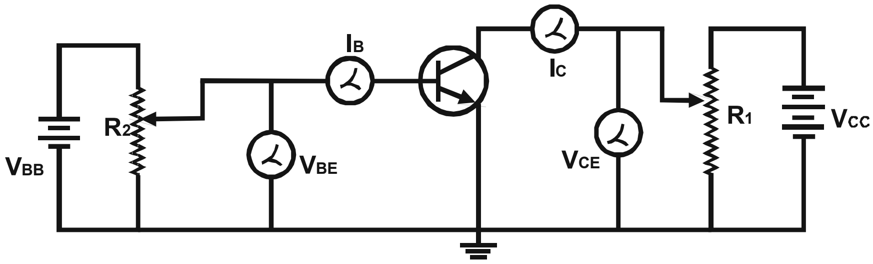



*Fig. 25.17*

**Common Emitter Static Characteristics**

***(a) Input Characteristic***

It shows how $I_B$ varies with changes in $V_{BE}$ when $V_{CE}$ is held constant at a particular value.

To begin with, voltage $V_{CE}$ is maintained constant at a convenient value and then $V_{BE}$ is increased in steps. Corresponding values of $I_B$ are noted at each step. The procedure is then repeated for a different but constant value of $V_{CE}$. A typical input characteristic is shown in Fig.25.18. Like CB connection, the overall shape resembles the forward characteristic of a P-N diode.

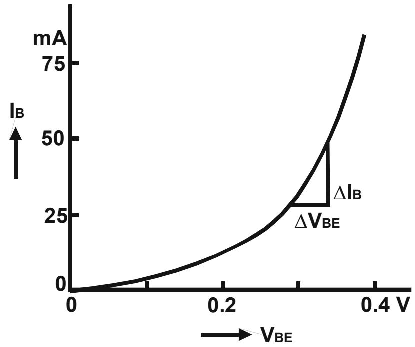



*Fig. 25.18*

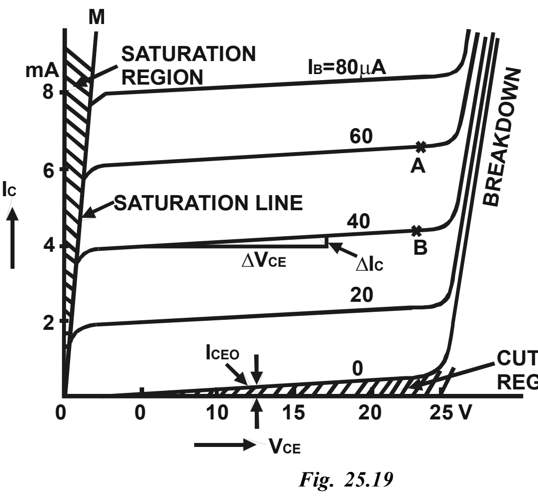



*Fig. 25.19*

The reciprocal of the slope gives the input resistance $R_{in}$ of the transistor.

$$R_{in} = [(1)/(\Delta I_B / \Delta V_{BE})] = (\Delta V_{BE} / \Delta I_B)$$

Due to initial non-linearity of the curve, $R_{in}$ varies considerably from a value of 4 kΩ near the origin to a value of 600 Ω over the more linear part of the curve.

***(b) Output or Collector Characteristic***

It indicates the way in which $I_C$ varies with changes in $V_{CE}$ when $I_B$ is held constant.

For obtaining this characteristic, first $I_B$ is set to a convenient value and maintained constant and then $V_{CE}$ is increased from zero in steps, $I_C$ being noted at each step. Next, $V_{CE}$ is reduced to zero and $I_B$ increased to another convenient value and the whole procedure repeated. In this way, a family of curves is obtained.

It is seen that as $V_{CE}$ increases from zero, $I_C$ rapidly increases to a near saturation level for a fixed value of $I_B$. As shown, a small amount of collector current flows even when $I_B = 0$. It is called $I_{CEO}$. Since main collector current is zero, the transistor is said to be cutoff.

It may be noted that if $V_{CE}$ is allowed to increase too far, C/B junction completely breaks down and due to this avalanche breakdown , $I_C$ increases rapidly and may cause damage to the transistor.

When $V_{CE}$ has very low value (ideally zero), the transistor is said to be saturated and it operates in the saturation region of the characteristic. Here, change in $I_B$ does not produce a corresponding change in $I_C$.

This characteristic can be used to find β.

$$\beta = \Delta I_C / \Delta I_B$$

We may select any two points A and B on the $I_B = 60$ μ A and 40μ A lines respectively and measure corresponding values of $I_C$ from the diagram for finding $\Delta I_c$. Since $\Delta I_B = (60-40) = 20$ μ A, β can be found.

This value of output resistance $R_{out} = \Delta V_{CE} / \Delta I_C$ over the near horizontal part of the characteristic varies from 10 kΩ to 50 kΩ.

***(c) Current Transfer Characteristic***

It indicates how $I_c$ varies with changes in $I_B$ when $V_{CE}$ is held constant at a given value.

Such a typical characteristic is shown in Fig. 25.20(a). Its slope gives $\beta = \Delta I_c / \Delta I_B$.

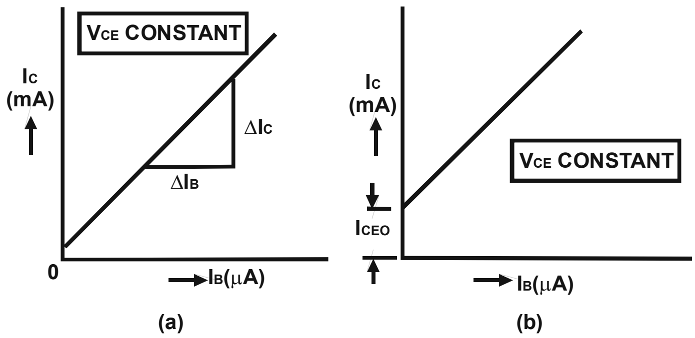



*Fig. 25.20*

From Fig. 25.20 (b), it is seen that a small collector current flows even when $I_B = 0$. It is the common-emitter leakage current $I_{CEO} = (1 + \beta) I_{co}$. Like $I_{CO}$, it is also due to the flow of minority carriers across the reverse-biased C/B junction.

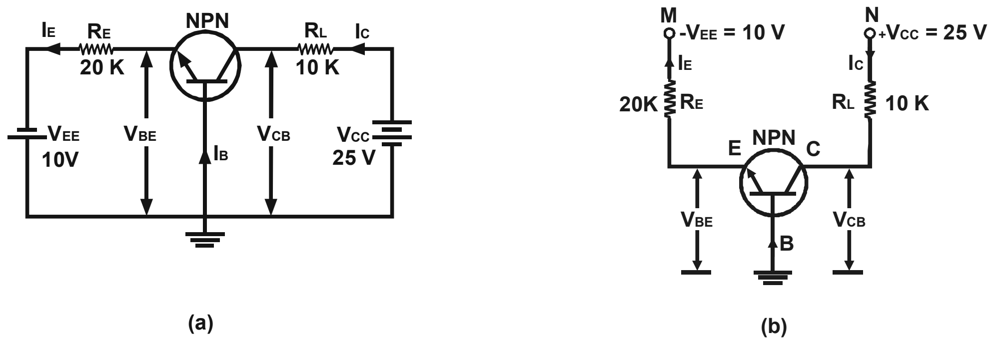



*Fig. 25.21*

**DIFFERENT WAYS OF DRAWING SCHEMATIC TRANSISTOR CIRCUITS**

In Fig. 25.21 (a) is shown a CB transistor circuit which derives its voltage and current requirements from two independent power source i.e. two different batteries. Correct battery connections can be done by remembering the transistor polarity rule that in an NPN transistor, both collector and base have to be POSITIVE with respect to the emitter. Of course, collector is a little more positive than base which means that between themselves, collector is at a slightly higher positive potential with respect to base. Conversely, base is at a little lower potential with respect to the collector.

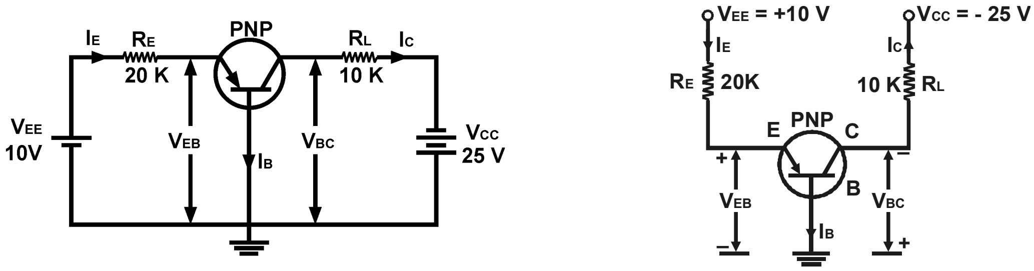



*Fig. 25.22*

Putting it in a slightly different way, we can say that collector is positive w.r.t. base and conversely, base is negative w.r.t. collector. That is why, potential difference between collector and base is written as $V_{CB}$ (and not $V_{BC}$) because terminal at higher potential is mentioned first. Same reasoning applies to $V_{BE}$. Fig. 25.21 (b) shows another and more popular way of indicating power supply voltage. Only one terminal of the battery is shown, the other terminal of each power supply is understood to be grounded (as is the base) even though not shown in the diagram.

Fig. 25.22 shows same circuit drawn for a PNP transistor. As expected, in this case, the battery polarities and directions of current flow are opposite to those shown in Fig. 25.23.

Fig. 25.23 (a) shows an NPN transistor connected in CE configuration with voltage and currents drawn two independent power sources.

As seen, battery connections and voltage markings are as per the rule. Fig. 25.23 (b) shown the more popular way of indicating power supply voltage.

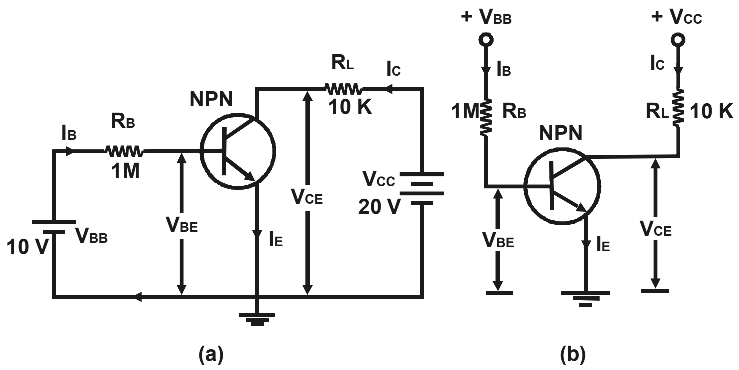



*Fig. 25.23*

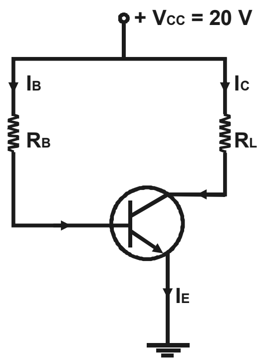



*Fig. 25.24*

As seen, both collector and base are positive with respect to the common electrode i.e. emitter. Hence, a single battery can be used to get proper voltage across the two as shown in Fig. 25.24

Fig. 25.25(a) shows the CC configuration of an NPN transistor and Fig.25.25 (b) shows the same circuit drawn differently.

**COMMON BASE FORMULAS**

Let us find the values of different voltage and currents for the circuit shown in Fig.25.21(b). Consider circuit MEBM. Starting from M, used applying Kirchhoff's voltage law, we get

$$+I_E R_E + V_{BE} - V_{EE} = 0$$

(a) or $I_E = [(V_{EE} - V_{BE}) / R_E)]$

where $V_{BE} = 0.3V$ ..... for Ge

$= 0.7\,V$ ...... for Si

Since generally, $V_{EE} > V_{BE}$, we can simplify the above to

$$I_E \cong V_{EE} / R_E$$

$$\cong 10V/20K = 0.5mA$$

Taking $V_{BE}$ into account and assuming silicon transistor,

$$I_E = (10 - 0.7)\,V/20\,K = 0.465\,mA$$

(b) $I_c = \alpha\, I_E \cong I_E$

neglecting leakage current $I_{CO}$

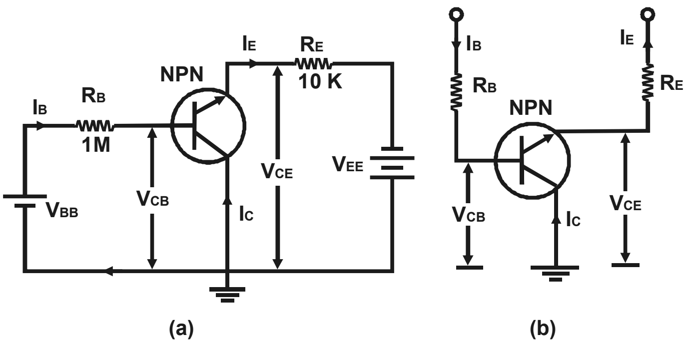



*Fig. 25.25*

(c) From circuit NCBN, we get

$$V_{CB} = V_{CC} - I_C R_L$$
$$\cong V_{CC} - I_E R_L \qquad \text{because} \quad I_C \cong I_E$$
$$V_{CB} = 25 - 0.5 \times 10 = 20V$$

**Common Emitter Formulas**

Consider the CE circuit of Fig. 25.26

Taking the emitter-base circuit, we have

$$I_B = [(V_{BB} - V_{BE})/(R_B)] \cong (V_{BB}/R_B)$$
$$I_C = \beta\, I_B \text{ - neglecting leakage current } I_{CEO}$$
$$V_{CE} = V_{CC} - I_C R_L$$

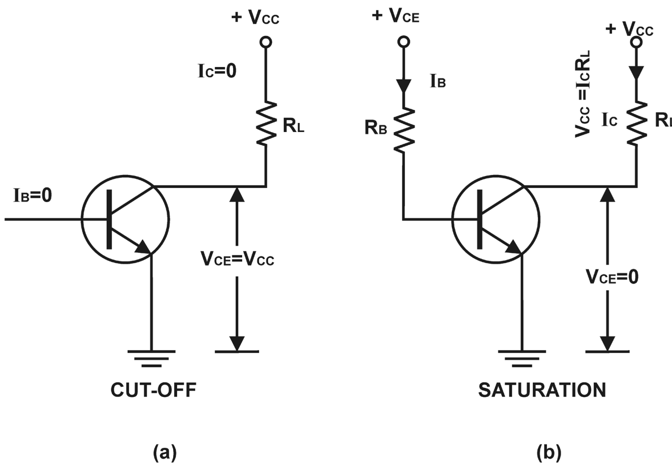



*Fig. 25.26*

**Cutoff and Saturation Points**

Consider the circuit of Fig. 25.26 (a). As we know

$$V_{CE} = V_{CC} - I_C R_L \qquad \text{...(i)}$$

Since $I_B = 0$, $I_C = 0$. Hence

$$V_{CE} = V_{CC}$$

Under these conditions, the transistor is said to cutoff for the simple reason that transistor does not conduct any current.* This value of $V_{CE}$ is written as $V_{CE\ (out\text{-}off)}$. Moreover, in cutoff, both the base-emitter and the base-collector junctions are reverse-biased.

If in Fig. 25.26(b), $I_B$ is increased by increasing $V_{BB}$, then $I_C$ is increased because $I_C = \beta\, I_B$. This increases drop across $R_L$ as a result of which $V_{CE}$ is decreased because $V_{CE} = V_{CC} - I_C R_L$. A certain value of $I_C$ is reached when $I_c R_L$ becomes equal to $V_{CC}$ itself.

In that case,

$$V_{CE} = V_{CC} - I_C R_L = V_{CC} - V_{CC} = 0$$

When $V_{CE} = 0$, the transistor is said to be in saturation because it then carries the maximum collector current called $I_{c(sat)}$. Even if $I_B$ is increased now, $I_C$ does not increase beyond its saturation value because under saturation condition, the relation $I_C = \beta I_B$ does not hold good. To summarize the above, we have that when a transistor is in saturation :

1. Whole of $V_{CC}$ drops across $R_L$
2. $I_C$ has maximum value of $I_{C(sat)} = V_{CC} / R_L$.

Normal operation of a transistor lies between the above two extreme conditions of out-off and saturation.

**Importance of $V_{CE}$**

The voltage $V_{CE}$ is very important in checking either the transistor is

(a) Defective
(b) Working in cutoff
(c) In saturation or well into saturation.

When $V_{CE} = V_{CC}$, the transistor is in cutoff i.e. it is turned OFF. When $V_{CE} = 0$, the transistor is in saturation i.e. it is turned full ON. When $V_{CE}$ is less than zero i.e. negative, the transistor is said to be well into saturation. In practice, both these conditions are avoided. For amplifier operation, $V_{CE} = ½\ V_{CC}$ i.e. transistor is operated at approximately ½ ON. In this way, variations in $I_B$ in either direction will control $I_C$ in both directions. In other words, when $I_B$ increases or decreases, $I_C$ also increases or decreases. However, if $I_B$ is OFF $I_C$ is also OFF. On the other hand, if collector has been turned fully ON, maximum $I_C$ flows. Hence, no further increase in $I_E$ can be reflected in $I_C$.

■■■

# CHAPTER : 26
# TRANSISTOR AMPLIFIERS

**CLASSIFICATION OF AMPLIFIERS**

Linear amplifiers are classified according to their mode of operation i.e. the way they operate according to a predetermined set of values. Various amplifier descriptions are based on the following factors :

1. **As based on its input**
   (a) Small-signal amplifier
   (b) Large-signal amplifier
2. **As based on its output**
   (a) Voltage amplifier
   (b) Power amplifier
3. **As based on its frequency response**
   (a) Audio-frequency (AF) amplifier
   (b) Intermediate-frequency (IF) amplifier
   (c) Radio-frequency (RF) amplifier
4. **As based on its biasing conditions**
   (a) Class-A (b) Class-AB
   (c) Class-B (d) Class-C
5. **As based on transistor configuration**
   (a) Common-Base (CB) Amplifier
   (b) Common-Emitter (CE) Amplifier
   (c) Common-Collector (CC) Amplifier

The description : small -signal, class-A, CE, voltage amplifier means that input signal is small, biasing condition is class-A, transistor configuration is common-emitter and its output concerns voltage amplification.

We will first take up the basic working of a single-stage amplifier i.e. an amplifier having one amplifying element connected in CB, CE and CC configuration.

**COMMON BASE (CB) AMPLIFIER**

Both Fig. 26.1 and 26.2 shows the circuit of a single-stage CB amplifier using NPN transistor. As seen, input ac signal is injected into the emitter-base circuit and output is taken from the collector-base circuit. The E/B junction is forward-biased by $V_{EE}$ whereas C/B junction is reverse-biased by $V_{CC}$. The Q-point or dc working conditions are determined by the batteries along with resistors $R_E$ and $R_C$.

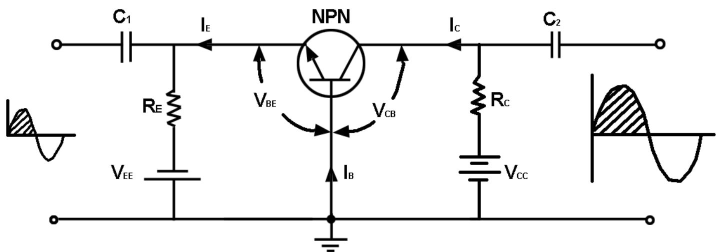



*Fig. 26.1*

When no signal is applied to the input circuit , the output just sits at the Q-point so that there is no output signal. Let us now see what happens when we apply an ac signal to the E/B junction via a coupling capacitor $C_1$ (which is assumed to offer no reactance to the signal).

**Circuit Operation**

When positive half-cycle of the signal is applied, then

1. Forward bias is decreased because $V_{BE}$ is already negative with respect to the ground as per biasing rule.
2. Consequently, $I_B$ is decreased.
3. $I_E$ and hence $I_C$ are decreased (because they are both nearly times the base current).
4. The drop $I_CR_C$ is decreased.
5. Hence, $V_{CB}$ is increased because

$$V_{CC} = V_{CB} + I_CR_C$$

or $$V_{CB} = V_{CC} - I_CR_C$$

It means that a positive output half-cycle is produced.

Since a positive-going input signal produces a positive-going output signal, there is no phase reversal between the two.

Voltage amplification in this circuit is possible by reason of relative input and output circuitry rather than current gain ($\alpha$) which is always less than unity. The input circuit has low resistance whereas output circuit has very large resistance. Although changes in the input and output currents are the same, the ac drop across $R_L$ is very large. Hence, changes in $V_{CB}$ (which is the output voltage) are much larger than changes in input ac signal. Hence, the voltage amplification.

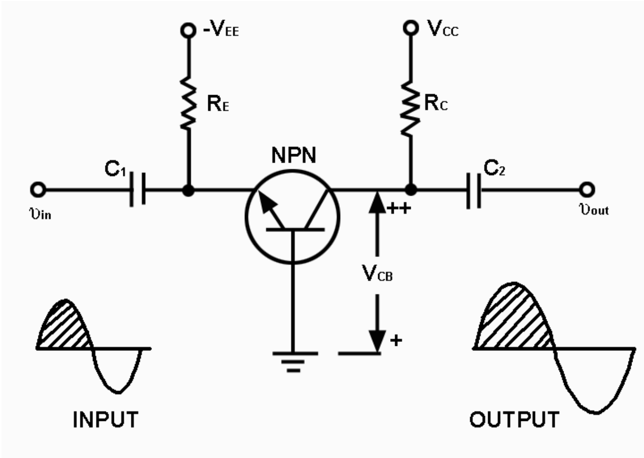



*Fig. 26.2*

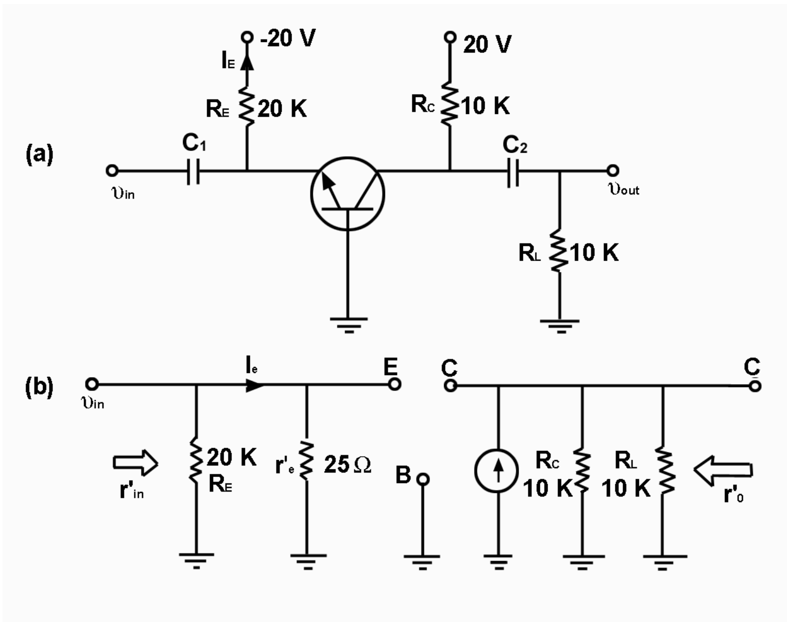



*Fig. 26.3*

**VARIOUS GAINS OF A CB AMPLIFIER**

**1. Input Resistance**

The ac input resistance of the transistor alone is given by the emitter junction resistance

$r_o' = 25\,mV/I_E$ or $50\,mV/I_E$

As seen from the ac equivalent circuit (Fig. 26.3 ), the input resistance of the stage is

$r_{in}' = r'_0 \parallel R_E$

**2. Output Resistance**

$r_0 = R_C$

If a load resistance $R_L$ is connected across output terminals, then output resistance of the stage is

$r_o' = R_L \parallel R_C$

**3. Current Gain**

$A_i = \alpha$

**4. Voltage Gain**

$A_v = (r_o/r_{in})$ for transistor alone

$= (r'_o/r'_{in})$ for the stage

**5. Power Gain**

$A_p = A_v . A_i$

The decibel gain is given by $G_p = 10 \log_{10} A_p$dB.

**Characteristics of a CB Amplifier**

Common-base amplifier has

1. very low input resistance (30-150 W)
2. very high output resistance (upto 500 K)
3. a current gain $\alpha < 1$
4. large voltage gain of about 1500
5. power gain of upto 30 dB
6. no phase reversal between input and output voltages.

**Uses**

One of the important uses of a CB amplifier is in matching a low-impedance circuit to a high-impedance circuit.

It also has high stability of collector current with temperature changes.

**COMMON EMITTER (CE) AMPLIFIER**

Fig. 26.4 and 26.5 shows the circuit of a single-stage CE amplifier using an NPN transistor. Here, base is the driven element. The input signal is injected into the base-emitter circuit. Whereas output signal is taken from the collector-emitter circuit. The E/B junction is forward-biased by $V_{BB}$ and C/B junction is reverse-biased by $V_{CC}$ (in fact, same battery $V_{cc}$ can provide dc power for both base and collector as in Fig. 26.5). The Q-point or working condition is determined by $V_{CC}$ together with $R_B$ and $R_{C.}$

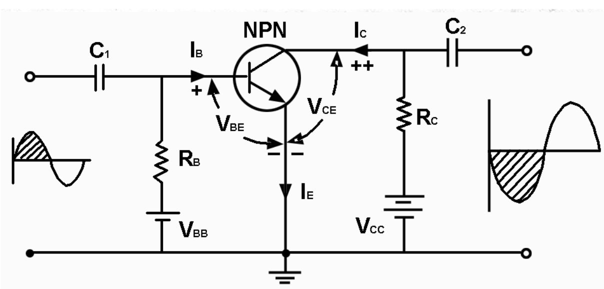



*Fig. 26.4*

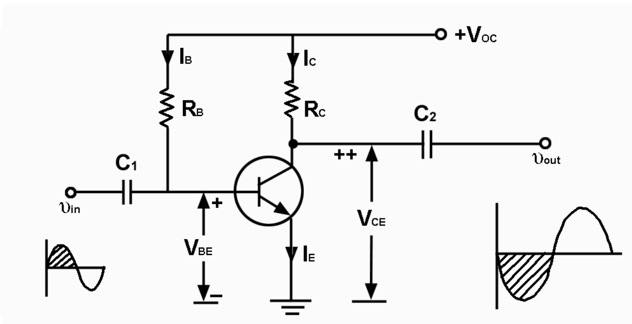



*Fig. 26.5*

Now, let us see what happens when an ac signal is applied at the input terminals of the circuit.

**Circuit Operation**

When positive half-cycles of the signal is applied.

1. $V_{BE}$ is increased because it is already positive with respect to the ground as per biasing rule.
2. It leads to increase in forward bias of the base-emitter junction.
3. $I_B$ is increased somewhat
4. $I_C$ is increased by $\beta$ times the increase in $I_B$.
5. Drop $I_C R_C$ is increased considerably and consequently
6. $V_{CE}$ which represents the output voltage is decreased

Now $\quad V_{CC} = V_{CE} + I_C R_C \quad$ or $\quad V_{CE} = V_{CC} - I_C R_C$

Hence, negative half-cycle of the output is obtained . It means that a positive-going input signal becomes a negative-going output signal as shown in Fig. 26.4 and 26.5.

**Various Gains of a CE Amplifier**

The ac equivalent of the given circuit (fig. 26.5) is similar to the one shown in Fig. 26.6 (b).

**1. Input Resistance**

When viewed from base, ac resistance of the emitter junctions is $\beta r_e$' As seen from Fig. 26.6 (b), circuit or stage input resistance is

$r'_{in} = R_B \quad \beta ' r_e$ remember $\beta$-rule

$\cong \beta ' r_e$ when $R_B >> \beta r_e$

It is called input resistance of the stage i.e. $r_{in(stage).}$

**2. Output Resistance**

$r_o = R_C$

However, if a load resistor $R_L$ is connected across the output terminals Fig. 26.6, then

$r'_o = R_c \parallel R_L = r_L$

**3. Voltage Gain**

$A_v = \beta\,(r_o'/\,r_{in}') \cong \beta.\,(\,r_o'/\beta\, r_e')$ If $R_B >> \beta r_e'$

it is the stage voltage gain

**4. Power Gain**

$A_P = A_V\,.\,Ai = \beta\; r_o'/r_e'$

$= G_p = 10\,\log_{10} A_P\, dB$

**5. Current Gain**

$A_i = \beta.$

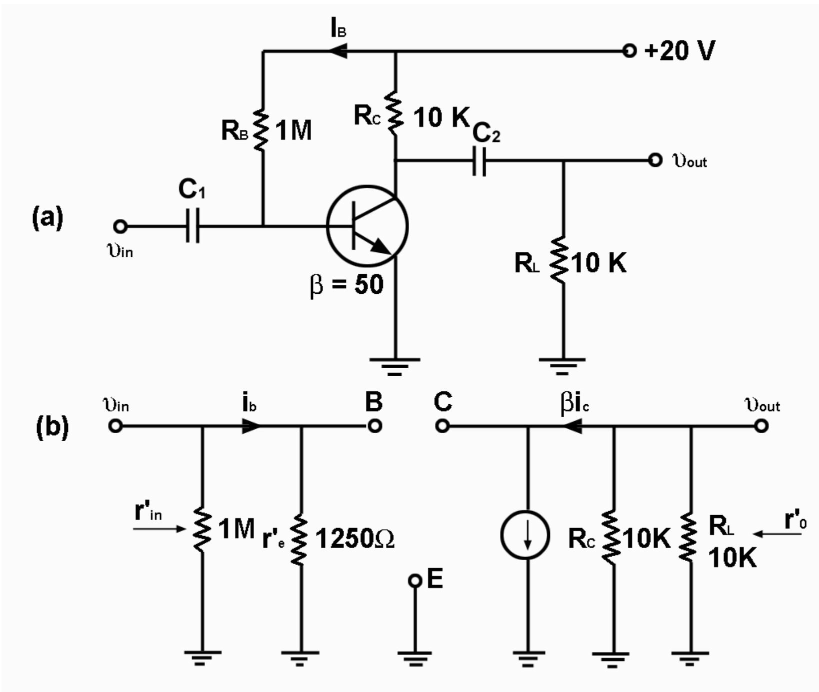



*Fig. 26.6*

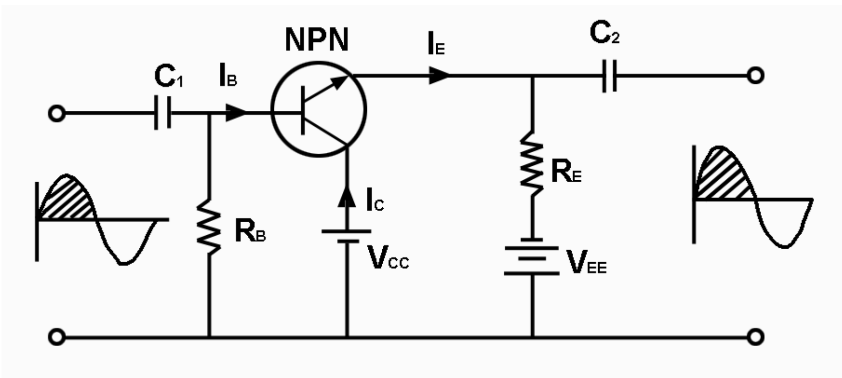



*Fig. 26.7*

**Characteristics of a CE Amplifier**

A CE transistor amplifier has the following characteristics:

1. It has moderately low input resistance ( 1 K to 2 K )
2. Its output resistance is moderately large ( 50 K or so )
3. Its current gain (β) is high ( 50-300)
4. It has very high voltage gain of the order of 1500 or so.
5. It produces very high power gain of the order of 10,000 or 40 dB.
6. It produces phase reversal of input signal i.e. input and output signals are 180° out of phase with each other.

**Uses**

Most of the transistor amplifiers are of CE type because of large gains in voltage, current and power. Moreover, its input and output impedance characteristics are suitable for many applications.

**Common Collector (CC) Amplifier**

Fig. 26. 7 & 26.8 show the circuit of a single-stage CC amplifier using an NPN transistor. The input signal is injected into

the base-collector circuit and output signal is taken out from the emitter-collector circuit. The E/B junction is forward-biased by $V_{EE}$ and C/B junction is reverse-biased by $V_{CC}$.

Let us now see what happens when an ac signal is applied across the input circuit.

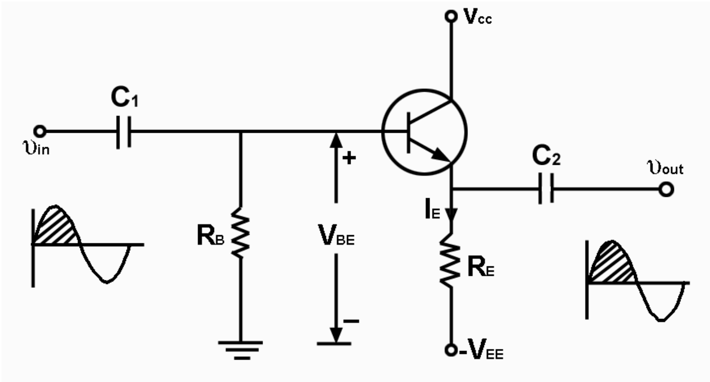



*Fig. 26.8*

**Circuit Operation**

When positive half-cycle of the signal is applied, then

1. Forward bias is increased since $V_{BE}$ is positive w.r.t. Collector i.e. ground
2. Base current is increased
3. Emitter current is increased
4. Drop across $R_E$ is increased
5. Hence, output voltage (i.e. drop across $R_E$) is increased.

Consequently, we get positive half-cycle of the output.

It means that a positive-going input signal results in a positive-going output signal and, consequently, the input and output signals are in phase with each other as shown in Fig. 26.8

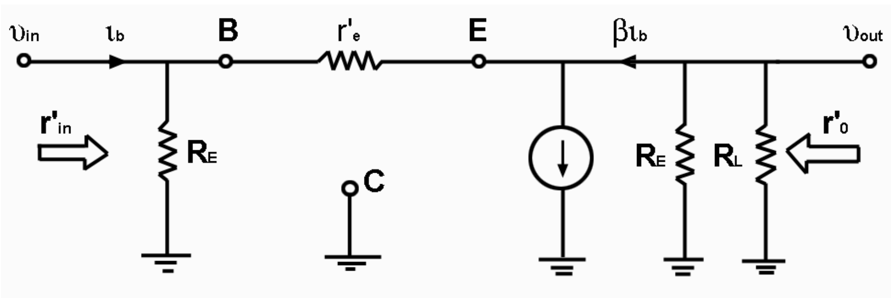



*Fig. 26.9*

**Various Gains of a CC Amplifier**

The ac equivalent circuit of the CC amplifier is given in Fig. 26.9

1. $r_{in}' = R_B \parallel \beta\ (r_e' + r_o')$
2. $r_o' = R_E \parallel\ R_L$
3. $A_i = 1 + \beta \cong \beta$
4. $A_v = [(r_o') / (r_o' + r_e')] \cong (r_o' / r_o') = 1$
   since usually $r_o' > r_e'$
5. $A_p = A_v . A_i$ and $G_p = 10 \log_{10} A_P$ dB

**Characteristics of a CC Amplifier**

A CC amplifier has the following characteristics:

1. High input impedance (20-500K)
2. Low output impedance (50Ω – 1 kΩ)
3. High current gain of (1+β) i.e. (50-500)
4. Voltage gain of less than 1
5. Power gain of 10 to 20 dB
6. No phase reversal of the input signal.

**Uses**

The CC amplifiers are used for the following purposes:

1. For impedance matching i.e. for connecting a circuit having high output impedance to one having low input impedance.
2. For circuit isolation
3. As a two-way amplifier since it can pass a signal in either direction
4. For switching circuits

**Phase Reversal in Amplifiers**

From the foregoing discussion of different amplifiers, it is seen that

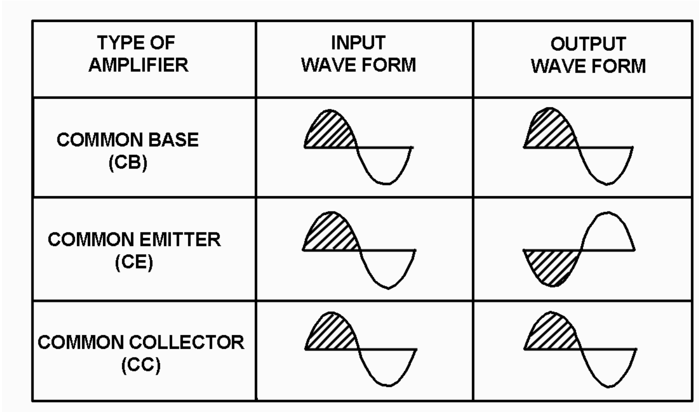



*Fig. 26.10*

1. CB amplifier does not change the phase of the input ac signal. As seen from Fig. 26.10, the input and output signals are in phase.
2. The CE amplifier inverts the input signal i.e. it causes a phase reversal of $180^0$ in the signal.
3. The CC amplifier does not change the phase of the input signal i.e. the input and output signals are in phase.

**Amplifier Classification Based on Biasing Conditions**

This classification is based on the amount of transistor bias and the amplitude of the input signal . It takes into account the portion of the cycle for which the transistor conducts. The three main classifications are : (i) Class A (ii) Class B and (iii) Class C

**Class-A Amplifier**

In this amplifier, the transistor is so biased that

1. Its Q-point is at the centre of the load line.
2. Amplitude of the input signal is such that the transistor operates over the linear portion of the load line.
3. Output current flows during the entire cycle of the input signal
4. Its conduction angle is $360^o$ as shown in Fig. 26.12

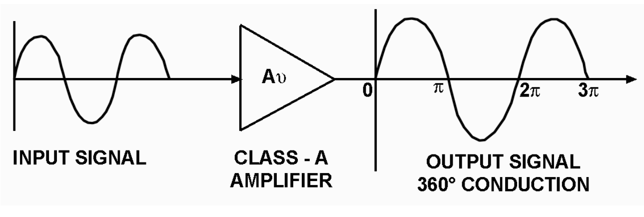



*Fig. 26.12*

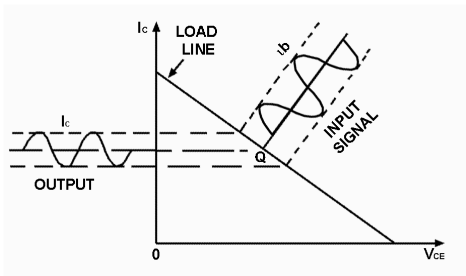



*Fig. 26.13*

The above points have been shown graphically in Fig. 26.13, where CE output characteristics have been taken. It is so because CE configuration is widely used in Electronics.

Here, $i_b$ represents the input signal whereas $i_e$ is the output signal. Since the transistor remains FR--biased throughout the input cycle, its output current flows for the entire cycle.

**Characteristics**

1. Since the transistor operates over the linear portion of the load line, the output waveform is exactly similar to the input waveform. Hence, class-A amplifiers are used where high-fidelity and distortion-free output is required as

in radio receivers and television sets.

2. Since its operation is restricted only over a small central region of the load line, this amplifier is meant only for amplifying input signals of small amplitude. Large signals will shift the Q-point into non-linear regions near saturation or cut-off and thus produce distortion.
3. Due to the limitation of the input signal amplitude, ac power output per active device (i.e. transistor) is small.
4. The overall efficiency of the amplifier circuit is

$$= \frac{\text{ac power delivered to the load}}{\text{total power delivered by dc supply}}$$

$$= \frac{\text{average ac power output}}{\text{average dc power input}}$$

The maximum possible overall efficiency of a class-A amplifier with resistive load is 25%.

5. The collector efficiency of a transistor is defined as

$$= \frac{\text{average ac power output}}{\text{average dc power input to transistor}}$$

The maximum possible collector efficiency of a class-A amplifier with resistive load is 50%

6. In case an output transformer is used, the maximum possible overall efficiency and maximum possible collector efficiency for a class-A amplifier are both 50%
7. Since under zero-signal condition, there is no ac output power, it means that all the power given to the transistor is wasted as heat. Hence, the transistor dissipates maximum power under zero-signal condition.

**Class-B Amplifier**

In this amplifier, the transistor is so biased that

1. The Q-point lies at the cut-off point
2. There is no output current when input current is zero
3. It conducts only for half-cycle of the input i.e. its conduction angle is only 180°(Fig. 26.13).

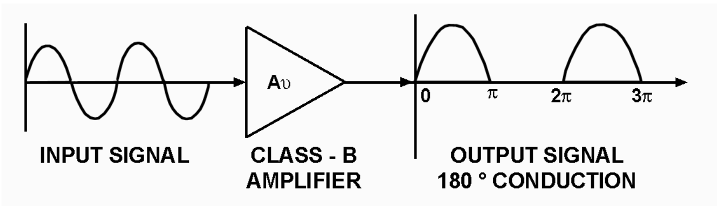



*Fig. 26.14*

The above facts have been shown graphically on the CE characteristics in Fig. 26.15. It is seen that output consists of only positive half-cycles, the negative half-cycles having been suppressed when transistor becomes biased below cut-off.

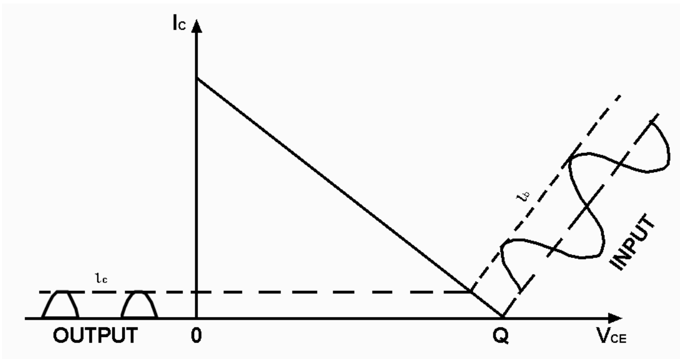



***Fig. 26.15***

**Characteristics**

1. Since negative half-cycles are totally absent from the output, the signal distortion is high as compared to class-A amplifiers. Hence, such amplifiers are not much used except for radio-frequency (RF) amplifiers.
2. Since input voltage is large, voltage amplification is reduced.
3 Zero-signal input represents worst case condition for class-A amplifiers but best condition for class-B amplifiers.
4. In class-B amplifiers, transistor dissipates more power as signal increases but opposite is the case in class-A amplifiers.
5. Average current in class-B operation is less than in class-A, hence power dissipated is less. Consequently,

maximum circuit (or overall) efficiency of a class-B amplifier is 78.5%

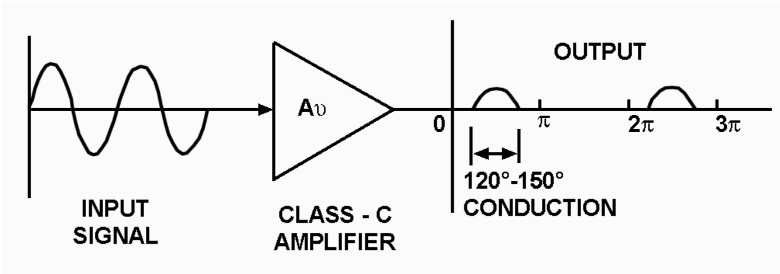



*Fig. 26.16*

**Class-C Amplifier**

In this amplifier, the transistor is biased much beyond cut-off (Fig. 26.17) Hence,

1. Output current flows only during a part of the positive half-cycle of the input signal. The conduction angle varies from 120°-150°. (Fig. 26.16)
2. There is no output current flow during any part of the negative half-cycle of the input signal.
3. Output signal has hardly any resemblance with the input signal. It consists of short pulses only.

The above facts have been shown on the $V_{CE}$ / $I_C$ characteristic of (Fig. 26.17).

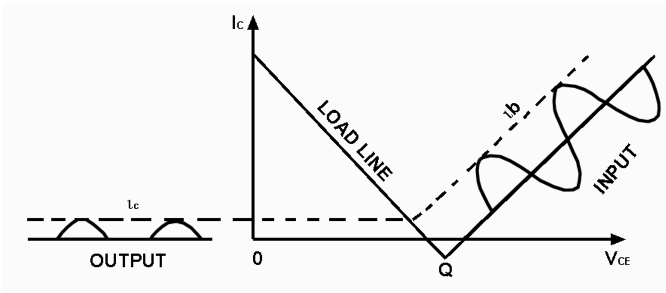



*Fig. 26.17*

**Characteristics**

1. Since output current flows for almost two-thirds of each half-cycle, power loss in such amplifiers is the least. Hence, they have very high efficiency of about 85- 90%.
2. Since output signal is much different from the input signal, class-C amplifiers suffer from very high distortion. Hence, such amplifiers are mainly used in oscillators for radio-frequency amplification where high efficiency is very essential regardless of distortion.

**Amplifier Coupling**

All amplifiers need some coupling network. Even a single-stage amplifier has to be coupled to the input and output devices. In the case of multistage systems, there is interstage coupling. The type of coupling used determines the characteristics of the cascaded amplifier. In fact, amplifiers are classified according to the coupling network used.

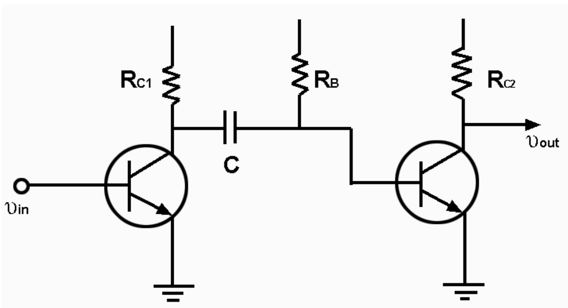



*Fig. 26.18*

The four methods of coupling are :-

**1. Resistance-capacitance (RC) Coupling**

It is also known as capacitive coupling and is shown in Fig. 26.18, Amplifiers using this coupling are known as RC-coupled amplifiers. Here, RC coupling network consists of two resistors $R_{C1}$ and $R_B$ and one capacitor C. The connecting link between the two stages is C. The function of RC coupling network is two fold:-

(a) To pass ac signal from one stage to the next.
(b) To block the passage of dc voltage from one stage to the next.

**2. Impedance Coupling or Inductive Coupling**

It is also known as choke-capacitance coupling and is shown in Fig. 26.19. Amplifiers using this coupling are known as impedance-coupled amplifiers. Here, the coupling network consists of $L_1$, C and network consists $R_B$. The impedance of the coupling coil depends on (i) its inductance and (ii) signal frequency.

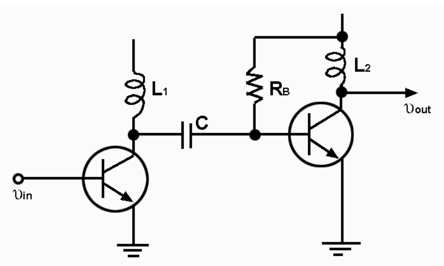



*Fig. 26.19*

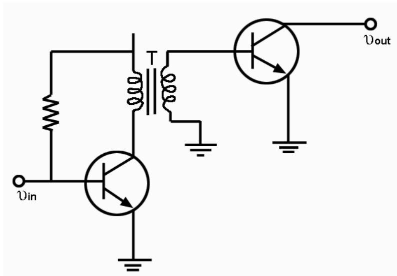



*Fig. 26.20*

**3. Transformer Coupling**

It is shown in Fig. 26.20. Since secondary of the coupling transformer conveys the ac component of the signal directly to the base of the second stage, there is no need for a coupling capacitor. Moreover, the secondary winding also provides a base return path, hence, there is no need for a base resistance. Amplifiers using this coupling are called transformer-coupled amplifiers.

**4. Direct Coupling**

It is shown in Fig. 26.21 This coupling is used where it is desirable to connect the load directly in series with the output terminal of the active circuit element. The examples of such load devices are (i) headphones (ii) loud-speakers (iii) dc meters (iv) dc relays and (v) input circuit of transistor etc. Of course, direct coupling is permissible only when

i) dc component of the output does not disturb the normal operation of the load device.

ii) Device resistance is so low that it does not appreciably reduce the voltage at the electrodes.

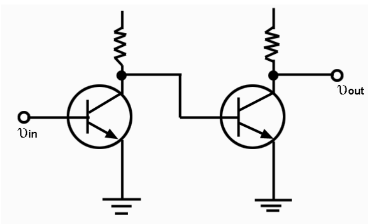



*Fig. 26.21*

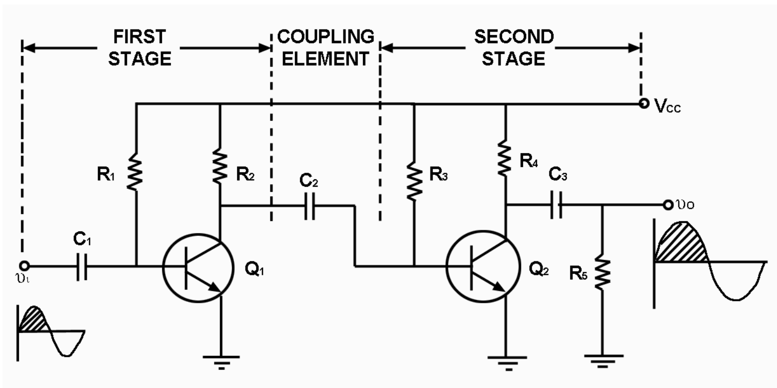



*Fig. 26.22*

**RC- coupled Two-stage Amplifier**

Fig. 26.22 shows a two-stage RC-coupled amplifier which consists of two single-stage transistor amplifiers using the CE configuration. The two resistors $R_2$ and $R_3$ and capacitor $C_2$ from the coupling network. $R_2$ is collector load of $Q_1$ and $R_4$ is that of $Q_2$. Capacitor $C_1$ couples the input signal whereas $C_3$ couples the output signal. $R_1$ and $R_3$ provide dc base bias. $R_5$ is the load resistor across $Q_2$.

**Circuit Operation**

The brief description of the circuit operation is as under :-

1. The input signal $v_i$ is amplified by $Q_1$. Its phase is reversed ( usual with CE connection).
2. The amplified output of $Q_1$ appears across $R_2$.
3. The output of the first stage across $R_2$ is coupled to the input of the second stage at $R_3$ by coupling capacitor $C_2$. This capacitor is also sometimes referred to as blocking capacitor because it blocks the passage of dc voltages and currents.
4. Thus the signal at the base of $Q_1$ is amplified by $Q_2$ and its phase is further reversed.
5. The ac output of $Q_2$ appears across $R_4$.
6. The output across $R_4$ is coupled by $C_3$ to load resistor $R_5$.
7. The output signal $v_o$ is the twice-amplified replica of the input signal $v_i$. It is in phase with $v_i$ because its phase has been reversed twice.

**Advantages of RC Coupling**

1. It requires no expensive or bulky components and no adjustments. Hence, it is small, light and inexpensive.

2. Its overall amplification is higher than that of the other couplings.
3. It has minimum possible nonlinear distortion because it does not use any coils or transformers which might pick up undesirable signals. Hence, there are no magnetic fields to interfere with the signal.
4. As shown in Fig. 26.23, it has a very flat frequency versus gain curve i.e. it gives uniform voltage amplification over a wide range from a few hertz to a few megahertz because resistor values are independent of frequency changes.

As seen from Fig. 26.22, amplifier gain falls off at very low as well as at very high frequencies, the fall in gain (called roll-off) is due to capacitive reactance of the coupling capacitor between the two stages. The high-frequency roll-off is due to output capacitance of the first stage, input capacitance of the second stage and the stray capacitance.

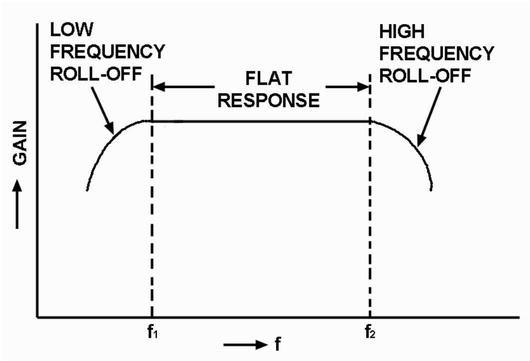



*Fig. 26.23*

The only drawback of this coupling is that due to large drop across collector load resistors, the collectors work at relatively small voltages unless higher supply voltage is used to overcome this large drop.

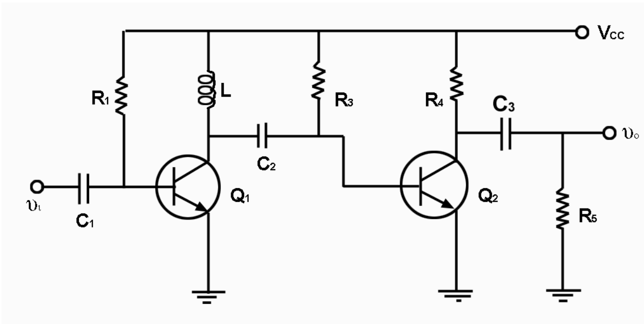



*Fig. 26.24*

**Impedance coupled Two-stage Amplifier**

The circuit is shown in Fig. 26.24 The coupling network consists of L $C_2$ and $R_3$ . The only basic difference between this circuit and the one shown in Fig. 26.22 is that inductor L has replaced the resistor $R_2$.

**Advantages and Disadvantages**

The biggest advantage of this coupling is that there is hardly any dc drop across L so that low collector supply voltages can be used.

However, it has many disadvantages:

1. It is larger, heavier and costlier than RC coupling.
2. In order to prevent the magnetic field of the coupling inductor from affecting the signal, the inductor turns are wound on a closed core and are also shielded.
3. Since inductor impedance depends on frequency, the frequency characteristics of this coupling are not as good. As those of RC coupling. The flat part of the frequency versus gain curve is small.

At low frequencies, the gain is low due to large capacitance offered by the coupling capacitor just as in RC-coupled amplifiers. The gain increases with frequency till it levels off at the middle frequencies of the audio range.

At relatively high frequencies, gain drops off again because of the increased reactance.

Hence, impedance coupling is rarely used beyond audio range.

**Transformer-coupled Two-stage Amplifier**

The circuit for such a cascaded amplifier is shown in Fig, 26.26. $T_1$ is the coupling transformer whereas $T_2$ is the output transformer. $C_1$ is the input coupling capacitor whereas $C_2$, $C_3$ and $C_4$ are the bypass capacitors. Resistors $R_1$ and $R_2$ as well as $R_4$ and $R_5$ form voltage-divider circuit whereas $R_3$ and $R_6$ are the emitter-stabilizing resistors.

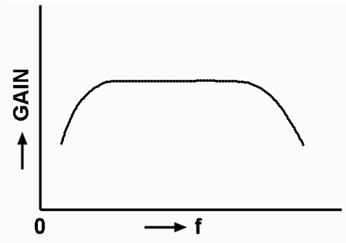



*Fig. 26.25*

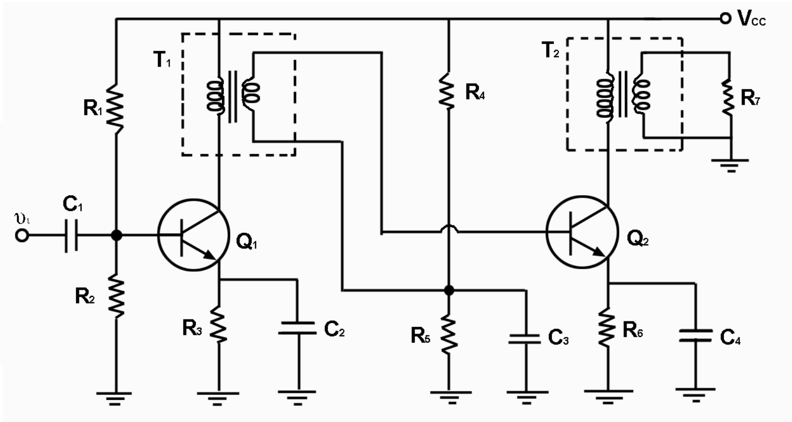



*Fig. 26.26*

**Circuit Operation**

When input signal is coupled through $C_1$ to the base of $Q_1$ , it appears in an amplified form in the primary of $T_1$. From there, it is passed on to the secondary by magnetic induction. Moreover, $T_1$ provides dc isolation between the input and output circuits. The secondary of $T_1$ applies the signal to the base of $Q_2$ from where it appears in an amplified form in the primary of $T_2$.

From there, it is passed on to the secondary by magnetic induction and finally appears across the matched load $R_7$.

**Advantages of Transformer Coupling**

1. The operation of a transformer-coupled system is basically more efficient because of low dc resistance of the primary connected in the collector circuit.
2. It provides a higher voltage gain.
3. It provides impedance matching between stages which is desirable for maximum power transfer. Typically, the input impedance of a transistor stage is less than its output impedance. Hence, secondary impedance of the interstage (or coupling ) transformer is typically lower than the primary impedance.

This coupling is effective when the final amplifier output is fed to a low-impedance load. For example, the impedance of a typical loud-speaker varies from 4 Ω to 16 Ω whereas output impedance of a transistor stage is several hundred ohms. Use of an output audio transformer can avoid the bad effects of such a mismatch.

**Disadvantages**

1. The coupling transformer is costly and bulky particularly when operated at audio frequencies because of its heavy iron core.
2. At radio frequencies, the inductance and winding capacitance present lot of problems.
3. It has poor frequency response because the transformer is frequency sensitive. Hence, the frequency range of transformer-coupled amplifiers is limited.
4. It tends to introduce 'hum' in the output.

**Direct-coupled Two-stage Amplifier**

These amplifiers operate without the use of frequency-sensitive components like capacitors, inductors and transformers etc. They are especially suited for amplifying:

1. Ac signals with frequencies as low as a fraction of a hertz.
2. Changes in dc voltages.

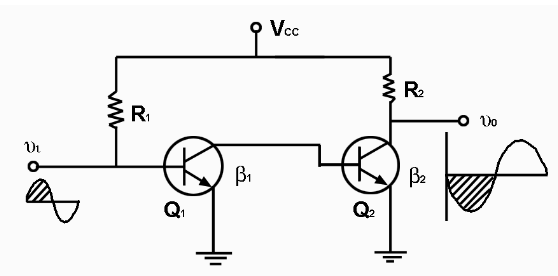



*Fig. 26.27*

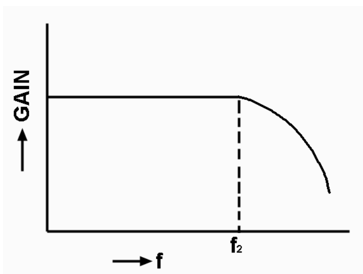



*Fig. 26.28*

Fig. 26.27 shows the circuit of such an amplifier which uses similar transistors each connected in the CE mode. Both stages employ direct coupling (i) collector of $Q_1$ connected directly to the base of $Q_2$ and (ii) load resistor $R_2$ is connected directly to the collector of $Q_2$. The resistor $R_1$ establishes the forward bias of $Q_1$ and also indirectly that of $Q_2$.

Any signal current at the base of $Q_1$ is amplified $\beta_1$ times and appears at collector of $Q_1$ and becomes base signal for $Q_2$. Hence, it is further amplified $\beta_2$ times, by $Q_2$. Obviously, signal current gain of the cascaded amplifier is

$$A_i = \beta_1 \times \beta_2 = \beta^2 \quad - \quad \text{if transistors are identical} \quad (\text{i.e. } \beta_1 = \beta_2)$$

**Advantages**
1. The circuit arrangement is very simple since it uses minimum number of components.
2. It is quite inexpensive.
3. It has the outstanding ability to amplify direct current (i.e. as dc amplifier ) and low-frequency signals.
4. It has no coupling or bypass capacitors to cause a drop in gain at low frequencies. As seen in Fig. 26.28, the frequency- response curve is flat upto upper cut-off frequency determined by stray wiring capacitance and internal transistor capacitance.

**Disadvantages**
1. It cannot amplify high-frequency signals.
2. It has poor temperature stability.

It is due to the fact that any variation in base current (due to temperature changes) in one stage is amplified in the following stage (or stages) thereby shifting the Q-point. However, stability can be improved by using emitter-stabilizing resistors.

**Applications**
Some of the applications of direct-coupled amplifiers are in :
1. Regulator circuits of electronic power supplies
2. Pulse amplifiers
3. Differential amplifiers
4. Computer circuitry
5. Electronic instruments

**Feedback Amplifiers**
A feedback amplifier is that in which a fraction of the amplified output is fed back to the input circuit. This partial dependence of amplifier input on its output helps to control the output. A feedback amplifier consists of two parts : an amplifier and a feedback circuit.

**i) Positive feedback**
If the feedback voltage (or current) is so applied as to increase the input voltage (i.e. it is in phase with it), then it is called positive feedback. Other names for it are : regenerative or direct feedback.

Since positive feedback produces excessive distortion, it is seldom used in amplifiers. However, because it increases the power of the original signal, it is used in oscillator circuits.

**ii) Negative feedback**
If the feedback voltage (or current) is so applied as to reduce the amplifier input (i.e. it is 180° out of phase with it), then it is called negative feedback. Other names for it are : degenerative or inverse feedback.

Negative feedback is frequently used in amplifier circuits.

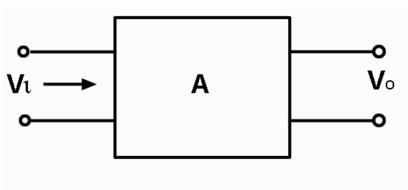


*Fig. 26.29*

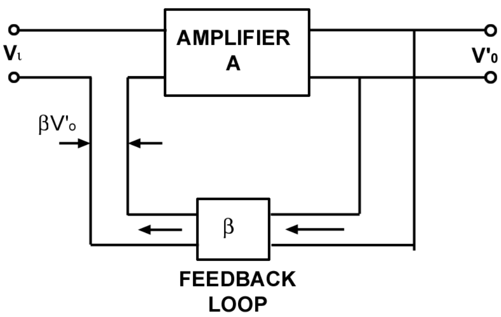


*Fig. 26.30*

**Principle of Feedback Amplifiers**
For an ordinary amplifier i.e. one without feedback, the voltage gain is given by the ratio of the output voltage $V_o$ and input voltage $V_i$. As shown in the block diagram of Fig. 26.29, the input voltage $V_i$ is amplified by a factor of A to the value $V_o$ of the output voltage.

$$\therefore \qquad A = (V_o / V_i)$$

The gain A is often called open-loop gain.

Suppose a feedback loop is added to the amplifier . If $V_o'$ is the output voltage with feedback, then a fraction $\beta$ of this voltage is applied to the input voltage which, therefore, becomes ( $V_i \pm \beta V_o'$) depending on whether the feedback voltage is in phase or antiphase with it. Assuming positive feedback, the input voltage will become ($V_i + \beta V_o'$). When amplified A times, it becomes $A(Vi + \beta V_o')$.

$\therefore\ A(V_i + \beta V_o') = V_o'$ or $V_o'(1-\beta A) = AV_i$

The amplifier gain A' with feedback is given by

$A' = (V_o'/V_i) = [(A)/(1-\beta A)]$ -from above
$A' = [(A)/(1-\beta A)]$ -positive feedback
and $A' = [(A)/(1-(-\beta A)]$
$= [(A)/(1+\beta A)]$ -negative feedback

The term '$\beta A$' is called feedback factor whereas $\beta$ is known as feedback ratio. The expression ( $1 \pm \beta A$ ) is called loop gain. The amplifier gain A' with feedback is also referred to as closed loop gain because it is the gain obtained after the feedback loop is closed.

**a) Negative Feedback**

$A' = [(A)/(1+\beta A)]$

Obviously, $A' < A$ because $|1+\beta A| > 1$. Suppose, A=90 and $\beta = 1/10 = 0.1$. Then, gain without feedback is 90 and with feedback is

$A' = [(A)/(1+\beta A)] = [(90)(1+0.1 \times 90)] = 9$

As seen, negative feedback reduces the amplifier gain. That is why it is called degenerative feedback. A lot of voltage gain is sacrificed due to negative feedback. When $|\beta A| >> 1$,

$A' \cong (A/\beta A) = (1/\beta)$

It means that A depends only on $\beta$. But it is very stable because it is not affected by changes in temperature, device parameters, supply voltage and from the ageing of circuit components etc. Since resistors can be selected very precisely with almost zero temperature-coefficient of resistance, it is possible to achieve a highly precise and stable gain with negative feedback.

**b) Positive Feedback**

The amplifier gain with positive feedback is given by

$A' = [(A)/(1-\beta A)]$

Since $|1-\beta A| < 1$, $A' > A$

Suppose gain without feedback is 90 and $\beta = (1/100 = 0.01)$, then gain with positive feedback is

$A' = [(90)/(1-(0.01 \times 90)] = 900$

Since positive feedback increases the amplifier gain, it is called regenerative feedback. If $\beta A = 1$, then mathematically, the gain becomes infinite which simply means that there is an output without any input. However, electrically speaking, this cannot happen. What actually happens is that the amplifier becomes an oscillator which supplies its own input. In fact, the two important and necessary conditions for circuit oscillation are:

1. The feedback must be positive.
2. Feedback factor must be unity i.e. $|\beta A| = 1$

**Advantages of Negative Feedback**

The numerous advantages of negative feedback outweigh its only disadvantage of reduced gain.
Among the advantages are:

1. Higher fidelity i.e. more linear operation.
2. Highly stabilized gain.
3. Increased bandwidth i.e. improved frequency response.
4. Less distortion.
5. Reduced noise.
6. Input and output impedances can be modified as desired.

■■■

# CHAPTER : 27
# DECIMAL PREFIXES

| Prefix | Symbol | Factor | |
|---|---|---|---|
| Tera | T | $10^{12}$ | = 1 000 000 000 000 |
| Giga | G | $10^{9}$ | = 1 000 000 000 |
| Mega | M | $10^{6}$ | = 1 000 000 |
| Kilo | K | $10^{3}$ | = 1 000 |
| Hecto | h | $10^{2}$ | = 100 |
| Deca | da | $10^{1}$ | = 10 |
| Deci | d | $10^{-1}$ | = 0.1 |
| Centi | c | $10^{-2}$ | = 0.01 |
| Milli | m | $10^{-2}$ | = 0.001 |
| Micro | μ | $10^{-6}$ | = 0.000 001 |
| Nano | n | $10^{-9}$ | = 0.000 000 001 |
| Pico | p | $10^{-12}$ | = 0.000 000 000 001 |
| Femto | F | $10^{-15}$ | = 0.000 000 000 000 001 |
| Atto | a | $10^{-18}$ | = 0.000 000 000 000 000 001 |

# CHAPTER : 28
# DECIMAL TO BINARY CONVERSION AND VICE -VERSA

## INTRODUCTION

We all are familiar with the number system in which an ordered set of ten symbols - 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9, known as digits - are used to specify any number. This number system is popularly known as the decimal number system. The radix or base of this number system is 10 (number of distinct digits). Any number is a collection of these digits. For example, 1982.365 signifies a number with an integer part equal to 1982 and a fractional part equal to 0.365, separated from the integer part with a Radix point (.) also known as decimal point. There are some other systems also used to represent numbers. Some of the other commonly used number systems are : binary, octal and hexadecimal number systems. These number system are widely used in digital systems like microprocessors, logic circuits, computers, etc. and therefore, the knowledge of these number systems is very essential for understanding, analysing and designing digital systems.

Computers and other digital circuits use binary signals but are required to handle data which may be numeric, alphabets or special characters. Therefore, the information available in any other form is required to be converted into suitable binary form before it can be processed by digital circuits. This means that the information available in the form of numerals, alphabets and special characters or in any combination of these must be converted into binary format. To achieve this, a process of coding is employed whereby each numeral, alphabet or special character is coded in a unique combination of 0s and 1s using a coding scheme, known as a code. The process of coding is known as encoding.

There can be a variety of coding schemes (codes) to serve different purposes, such as arithmetic operations, data entry, error detection and correction, etc. In digital systems, a large number of codes are in use. Selection of a particular code depends on its suitability for the purpose. In one digital system, different codes may be used for different operations and it may be necessary to convert data from one code to another code. For this purpose, code converter circuits are required which will be discussed later.

## NUMBER SYSTEMS

In general, in any number system there is an ordered set of symbols known as digits with rules defined for performing arithmetic operations like addition, multiplication, etc. A collection of these digits makes a number which in general has two parts - integer and fractional, set apart by a radix point (.), that is

$$(N)_b = \underbrace{d_{n-1}d_{n-2}\ldots d_i d_0}_{\text{Integer portion}} \underset{\substack{\uparrow \\ \text{Radix} \\ \text{Point}}}{\cdot} \underbrace{d_{-1}d_{-2}\ldots d_{-f}\ldots d_{-m}}_{\text{fractional Portion}} \qquad (1)$$

where, N = a number

b = radix or base of the number system

n = number of digits in integer portion

m = number of digits in fractional portion

$d_{n-1}$ = most significant digit (msd)

$d_{-m}$ = least significant digit (lsd)

and $0 \le (d_i \text{ or } d_{-f}) \le b-1$

The digits in a number are placed side by side and each position in the number is assigned a weight or index of importance by some predesigned rule. Table 28.1 gives the details of commonly used number systems.

**Table 28.1 Characteristics of commonly used number systems**

| Number system | Base or radix (b) | Symbols used $(d_i \text{ or } d_{-f})$ | Weight assigned to position: i | -f | Example |
|---|---|---|---|---|---|
| Binary | 2 | 0, 1 | $2^i$ | $2^{-f}$ | 1011.11 |
| Octal | 8 | 0, 1, 2, 3, 4, 5, 6, 7 | $8^i$ | $8^{-f}$ | 3567.25 |
| Decimal | 10 | 0, 1, 2, 3, 4, 5, 6, 7 8, 9 | $10^i$ | $10^{-f}$ | 3974.57 |
| Hexadecimal | 16 | 0, 1, 2, 3, 4, 5, 6, 7 8, 9, A, B, C, D, E, F | $16^i$ | $16^{-f}$ | 3FA9.56 |

**BINARY NUMBER SYSTEM**

The number system with base (or radix) two is known as the binary number system. Only two symbols are used to represent numbers in this system and these are 0 and 1. These are known as bits. This system has the minimum base (0 is not possible and 1 is not useful). It is a positional system, that is every position is assigned a specific weight.

Table 28.2 illustrates counting in binary number system. The corresponding decimal numbers are given in the right-hand column. Similar to decimal number system, the left-most bit is known as the most significant bit (MSB) and the right-most bit is known as the least significant bit (LSB). Any number of 0s can be added to the left of the number without changing the value of the number. In the binary number system, a group of four bits is known as a nibble, and a group of eight bits is known as a byte.

**Table 28.2, 4-bit binary numbers and their corresponding decimal numbers**

| Binary Number | | | | Decimal Number | |
|---|---|---|---|---|---|
| $B_3$ | $B_2$ | $B_1$ | $B_0$ | $D_1$ | $D_0$ |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 1 |
| 0 | 0 | 1 | 0 | 0 | 2 |
| 0 | 0 | 1 | 1 | 0 | 3 |
| 0 | 1 | 0 | 0 | 0 | 4 |
| 0 | 1 | 0 | 1 | 0 | 5 |
| 0 | 1 | 1 | 0 | 0 | 6 |
| 0 | 1 | 1 | 1 | 0 | 7 |
| 1 | 0 | 0 | 0 | 0 | 8 |
| 1 | 0 | 0 | 1 | 0 | 9 |
| 1 | 0 | 1 | 0 | 1 | 0 |
| 1 | 0 | 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | 0 | 1 | 2 |
| 1 | 1 | 0 | 1 | 1 | 3 |
| 1 | 1 | 1 | 0 | 1 | 4 |
| 1 | 1 | 1 | 1 | 1 | 5 |

**Binary-to-Decimal Conversion**

Any binary number can be converted into its equivalent decimal number using the weights assigned to each bit position as given in Table 28.1.

***Example 1*** Find the decimal equivalent of the binary number $(1\,1\,1\,1\,1)_2$.

***Solution*** The equivalent decimal number is

$$= 1\times 2^4 + 1\times 2^3 + 1\times 2^2 + 1\times 2^1 + 1\times 2^0$$
$$= 16 + 8 + 4 + 2 + 1$$
$$= (31)_{10}$$

To differentiate between numbers represented in different number systems, either the corresponding number system may be specified along with the number or a small subscript at the end of the number may be added signifying the number system. For example, $(1000)_2$ represents a binary number and is not one thousand.

***Example 2*** Determine the decimal numbers represented by the following binary numbers :

(a) 110101 (b) 101101 (c) 11111111 (d) 00000000

***Solution***

(a) $$(110101)_2 = 1\times 2^5 + 1\times 2^4 + 0\times 2^2 + 0\times 2^2 + 0\times 2^1 + 1\times 2^0$$
$$= 32 + 16 + 0 + 4 + 0 + 1$$
$$= (53)_{10}$$

(b) $$(101101)_2 = 32 + 0 + 8 + 4 + 0 + 1$$
$$= (45)_{10}$$

(c) $$(11111111)_2 = 128+64+32+16+8+4+2+1$$
$$= (255)_{10}$$

(d) $$(00000000)_2 = (0)_{10}$$

***Example 3.*** Determine the decimal numbers represented by the following binary numbers :
(a) 101101.10101 (b) 1100.1011 (c) 1001.0101 (d) 0.10101

***Solution***

(a) $$(101101.10101)_2 = 1\times 2^5 + 0\times 2^4 + 1\times 2^3 + 1\times 2^2 + 0\times 2^1 + 1\times 2^0$$
$$+1\times 2^{-1} + 0\times 2^{-2} + 1\times 2^{-3} + 0\times 2^{-4} + 1\times 2^{-5}$$
$$= 32+0+8+4+0+1+\frac{1}{2}+0+\frac{1}{8}+0+\frac{1}{32}$$
$$= (45.65625)_{10}$$

(b) $$(1100.1011)_2 = 8+4+0+0+0.5+0+0.125+0.0625$$
$$= (12.6875)_{10}$$

(c) $$(1001.0101)_2 = 8+0+0+1+00.25+0+0.0625$$
$$= (9.3125)_{10}$$

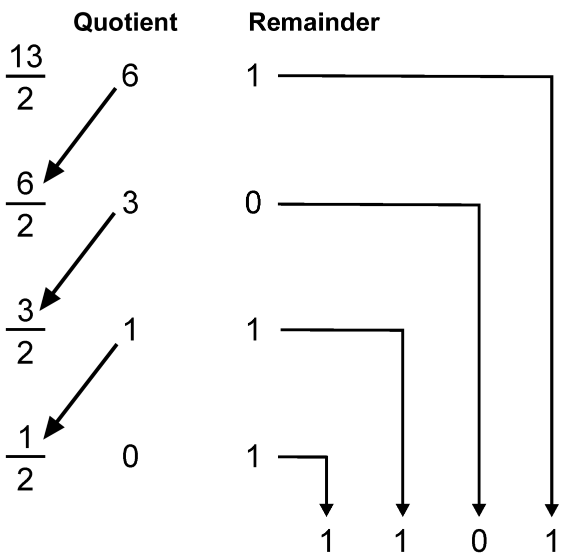
(d) $$(0.10101)_2 = 0.5+0+0.125+0+0.03125$$
$$= (0.65625)_{10}$$

**Decimal-to-Binary Conversion**

Any decimal number can be converted into its equivalent binary number. For integers, the conversion is obtained by continuous division by 2 and keeping track of the remainder, while for fractional parts, the conversion is affected by continuous multiplication by 2 and keeping track of the integers generated. The conversion process is illustrated by the following examples.

***Example 4*** Convert $(13)_{10}$ to an equivalent base-2 number.

***Solution***



***Example 5*** Convert $(0.65625)_{10}$ to an equivalent base -2 number.

Solution

$$\begin{array}{ccccc} 0.65625 & 0.31250 & 0.62500 & 0.25000 & 0.50000 \\ \underline{\times 2} & \underline{\times 2} & \underline{\times 2} & \underline{\times 2} & \underline{\times 2} \\ 1.31250 & 0.62500 & 1.25000 & 0.50000 & 1.00000 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 0 & 1 & 0 & 1 \end{array}$$

Thus, $(0.65625)_{10} = (0.10101)_2$

***Example 6*** Express the following decimal numbers in the binary form :
(a) 25.5 (b) 10.625 (c) 0.6875

***Solution***

(a) Integer Part

| | Quotient | Remainder |
|---|---|---|
| $\frac{25}{2}$ | 12 | 1 |
| $\frac{12}{2}$ | 6 | 0 |
| $\frac{6}{2}$ | 3 | 0 |
| $\frac{3}{2}$ | 1 | 1 |
| $\frac{1}{2}$ | 0 | 1 |

1 1 0 0 1

Therefore, $(25)_{10} = (11001)_2$

Fractional part

$$\begin{array}{c} 0.5 \\ \underline{\times 2} \\ 1.0 \\ \downarrow \\ 1 \end{array}$$

i.e., $(0.5)_{10} = (0.1)_2$

Therefore, $(25.5)_{10} = (11001.1)_2$

(b) Integer part $(10)_{10} = (1010)_2$

Fractional part

$$\begin{array}{ccc} 0.625 & 0.250 & 0.500 \\ \underline{\times 2} & \underline{\times 2} & \underline{\times 2} \\ 1.250 & 0.500 & 1.000 \\ \downarrow & \downarrow & \downarrow \\ 1 & 0 & 1 \end{array}$$

i.e., $(0.625)_{10} = (101)_2$

Therefore, $(10.625)_{10} = (1010.101)_2$

(c)

$$
\begin{array}{cccc}
0.6875 & 0.3750 & 0.7500 & 0.5000 \\
\underline{\quad\times 2} & \underline{\quad\times 2} & \underline{\quad\times 2} & \underline{\quad\times 2} \\
1.3750 & 0.7500 & 1.5000 & 1.0000 \\
\downarrow & \downarrow & \downarrow & \downarrow \\
1 & 0 & 1 & 1
\end{array}
$$

Therefore, $(0.6875)_{10} = (0.1011)_2$

## SIGNED BINARY NUMBERS

### Sign-Magnitude Representation

In the decimal number system a plus (+) sign is used to denote a positive number and a minus (-) sign for denoting a negative number. The plus sign is usually dropped, and the absence of any sign means that the number has positive value. This representation of numbers is known as signed number. As is well known, digital circuits can understand only two symbols, 0 and 1; therefore, we must use the same symbols to indicate the sign of the number also. Normally, an additional bit is used as the sign bit and it is placed as the most significant bit. A 0 is used to represent a positive number and a 1 to represent a negative number. For example, an 8-bit signed number 01000100 represents a positive number and its value ( magnitude) is $(1000100)_2 = (68)_{10.}$ The left most 0 (MSB) indicates that the number is positive. On the other hand, in the signed binary form, 11000100 represents a negative number with magnitude $(1000100)_2 = (68)_{10}$. The 1 in the left most position (MSB) indicates that the number is negative and the other seven bits give its magnitude. This kind of representation for signed numbers is known as sign-magnitude representation. The user must take care to see the representation used while dealing with the binary numbers.

***Example 7*** Find the decimal equivalent of the following binary numbers assuming sign-magnitude representation of the binary numbers.

(a) 101100 (b) 001000 (c) 0111 (d) 1111

***Solution***

(a) Sign bit is 1, which means the number is negative.

Magnitude = 01100 = $(12)_{10}$

$\therefore \quad (101100)_2 = (-12)_{10}$

(b) Sign bit is 0, which means the number is positive.

Magnitude = 01000 = 8

$\therefore \quad (001000)_2 = (+8)_{10}$

(c) $(0111)_2 = (+7)_2$

(d) $(1111)_2 = (-7)_2$

### One's Complement Representation

In a binary number, if each 1 is replaced by 0 and each 0 by 1, the resulting number is known as the one's complement of the first number. In fact, both the numbers are complement of each other. If one of these numbers is positive, then the other number will be negative with the same magnitude and vice-versa. For example, $(0101)_2$ represents $(+5)_{10}$, whereas $(-5)_{10}$ in this representation. This method is widely used for representing signed numbers. In this representation also, MSB is 0 for positive numbers and 1 for negative numbers.

***Example 8*** Find the one's complement of the following binary numbers.

(a) 0100111001 (b) 11011010

***Solution***

(a) 1011000110 (b) 00100101

***Example 9*** Represent the following numbers in one's complement form.

(a) +7 and -7 (b) +8 and -8 (c) + 15 and - 15

Solution In one's complement representation,

(a) $(+7)_{10} = (0111)_2$

and $(-7)_{10} = (1000)_2$

(b) $(+8)_{10} = (01000)_2$

and $(-8)_{10} = (10111)_2$

(c) $(+15)_{10} = (01111)_2$

and $(-15)_{10} = (10000)_2$

From the above examples, it can be observed that for an n-bit number, the maximum positive number which can be represented in 1's complement representation is $(2^{n-1} - 1)$ and the maximum negative number is - $(2^{n-1} - 1)$.

**Two's Complement Representation**

If 1 is added to 1's complement of a binary number, the resulting number is known as the two's complement of the binary number. For example, 2's complement of 0101 is 1011. Since 0101 represents $(+5)_{10}$, therefore, 1011 represents $(-5)_{10}$ in 2's complement representation. In this representation also, if the MSB is 0 the number is positive, whereas if the MSB is 1 the number is negative. For an n-bit number, the maximum positive number which can be represented in 2's complement form is $(2^{n-1} - 1)$ and the maximum negative number is $-2^{n-1}$. Table 2.3 gives sign-magnitude, 1's and 2's complement numbers represented by 4-bit binary numbers. From the table, it is observed that the maximum positive number is 0111 =+7, whereas the maximum negative number is 1000 = -8 using four bits in 2's complement format.

It is also observed that the 2's complement of the 2's complement of a number is the number itself.

**Table 28.3 Sign-magnitude, 1's and 2's complement representation using four bits**

| Decimal number | Binary number: Sign-magnitude | One's complement | Two's complement |
|---|---|---|---|
| 0 | 0000 | 0000 | 0000 |
| 1 | 0001 | 0001 | 0001 |
| 2 | 0010 | 0010 | 0010 |
| 3 | 0011 | 0011 | 0011 |
| 4 | 0100 | 0100 | 0100 |
| 5 | 0101 | 0101 | 0101 |
| 6 | 0110 | 0110 | 0110 |
| 7 | 0111 | 0111 | 0111 |
| -8 | - | - | 1000 |
| -7 | 1111 | 1000 | 1001 |
| -6 | 1110 | 1001 | 1010 |
| -5 | 1101 | 1010 | 1011 |
| -4 | 1100 | 1011 | 1100 |
| -3 | 1011 | 1100 | 1101 |
| -2 | 1010 | 1101 | 1110 |
| -1 | 1001 | 1110 | 1111 |
| -0 | 1000 | 1111 | - |

***Example 10.*** Find the 2's complement of the numbers :
(i) 01001110 (ii) 00110101

***Solution***

| | | |
|---|---|---|
| (i) | Number | 0 1 0 0 1 1 1 0 |
| | 1's complement | 1 0 1 1 0 0 0 1 |
| | Add 1 | 1 |
| | | 1 0 1 1 0 0 1 0 |

| | | |
|---|---|---|
| (ii) | Number | 0 0 1 1 0 1 0 1 |
| | 1's complement | 1 1 0 0 1 0 1 0 |
| | Add 1 | 1 |
| | | 1 1 0 0 1 0 1 1 |

From the above example, we observe the following :

1. If the LSB of the number is 1, its 2's complement is obtained by changing each 0 to 1 and 1 to 0 except the least-significant bit.
2. If the LSB of the number is 0, its 2's complement is obtained by scanning the number from the LSB to MSB bit by bit and retaining the bits as they are up to and including the occurrence of the first 1 and complement all other bits.

***Example 11.*** Find two's complement of the numbers :
(i) 01100100 (ii) 10010010 (iii) 11011000 (iv) 01100111]

***Solution*** Using the rules of conversion given above, we obtain

| | | |
|---|---|---|
| | | ↓ ↓ ↓ |
| (i) | Number | 0 1 1 0 0 1 0 0 |
| | 2's Complement | 1 0 0 1 1 1 0 0 |

| | | |
|---|---|---|
| | | ↓ ↓ |
| (ii) | Number | 1 0 0 1 0 0 1 0 |
| | 2's Complement | 0 1 1 0 1 1 1 0 |

| | | |
|---|---|---|
| | | ↓ ↓ ↓ ↓ |
| (iii) | Number | 1 1 0 1 1 0 0 0 |
| | 2's Complement | 0 0 1 0 1 0 0 0 |

| | | |
|---|---|---|
| | | ↓ |
| (iv) | Number | 0 1 1 0 0 1 1 1 |
| | 2's Complement | 1 0 0 1 1 0 0 1 |

***Example 12***. Represent $(-17)_{10}$ in

(i) Sign-magnitude
(ii) one's complement,
(iii) two's complement representation

***Solution*** The minimum number of bits required to represent $(+17)_{10}$ in signed number format is six.

$$\therefore \quad (+17)_{10} = (010001)_2$$

Therefore, $(-17)_{10}$ is represented by

(i) 110001 in sign-magnitude representation
(ii) 101110 in 1's complement representation
(iii) 101111 in 2's complement representation.

**BINARY ARITHMETIC**

We all are familiar with the arithmetic operations such as addition, subtraction, multiplication, and division of decimal numbers. Similar operations can be performed on binary numbers ; Infact, binary arithmetic is much simpler than decimal arithmetic because here only two digits, 0 and 1 are involved.

**Binary Addition**

The rules of binary addition are given in Table 28.4

**Table 28.4 Rules of Binary Addition**

| *Augend* | *Addend* | *Sum* | *Carry* | *Result* |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 |
| 1 | 1 | 0 | 1 | 10 |

In the first three rows above, there is no carry, that is, carry = 0, whereas in the fourth row a carry is produced (since the largest digit possible is 1), that is, carry = 1, and similar to decimal addition it is added to the next higher binary position.

***Example 13.*** Add the binary numbers :

(i) 1011 and 1100 (ii) 0101 and 1111

***Solution***

(i)

$$\begin{array}{ccccc} & 1 & 0 & 1 & 1 \\ (+) & 1 & 1 & 0 & 0 \\ \hline 1 & 0 & 1 & 1 & 1 \\ \uparrow & & & & \\ \text{carry} & & & & \end{array}$$

(ii)

$$\begin{array}{cccccl} & (1) & (1) & (1) & \leftarrow & \text{carry} \\ & 0 & 1 & 0 & 1 & \\ (+) & 1 & 1 & 1 & 1 & \\ \hline 1 & 0 & 1 & 0 & 0 & \\ \uparrow & & & & & \\ \text{carry} & & & & & \end{array}$$

***Example 14.*** Add the binary numbers :

$$\begin{array}{cccccccc} 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline \end{array}$$

***Solution***

$$\begin{array}{cccccccccl} & & & & \text{Two pair of 1's in the previous column} & & & & & \\ & (1) & (1) & (1) & \begin{pmatrix}1\\1\end{pmatrix} & (1) & (1) & (1) & \leftarrow & \text{one pair of 1's in the previous column} \\ & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & \\ & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \\ & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \\ & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \\ \hline 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & \\ \text{carry} & & & & & \uparrow & \uparrow & \uparrow & & \text{Even number of 1's in column} \\ & \uparrow & \uparrow & \uparrow & \uparrow & & & & & \text{odd number of 1's in column} \end{array}$$

$\therefore$ The sum = 1 1 1 1 1 0 0 1 0

From the above example, we observe the following :

(i) If the number of 1's to be added in a column is even then the sum bit is 0, and if the number of 1's to be added in a column is odd then the sum bit is 1.

(ii) Every pair of 1's in a column produces a carry (1) to be added to the next higher bit column.

**Binary Substraction**

The rule of binary subtraction are given in Table 28.5

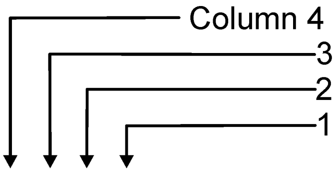
**Table 28.5, Rules of binary subtraction**

| Minuend | Subtrahend | Difference | Borrow |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 1 |
| 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 |

Except in the second row above, the *borrow = 0*. When the *borrow* = 1, as in the second row, this is to be subtracted from the next higher binary bit as it is done in decimal subtraction.

***Example 15.*** Perform the following subtraction :

$$\begin{array}{r} 1\;0\;1\;1 \\ -\;0\;1\;1\;0 \\ \hline \end{array}$$




*Solution*

$$\begin{array}{rl} 1\;0\;1\;1 & \text{Minuend} \\ (-)\,0\;1\;1\;0 & \text{Subtrahend} \\ \hline 0\;1\;0\;1 & \text{Difference} \\ \hline \end{array}$$

**Binary Multiplication**

Binary multiplication is similar to decimal multiplication. In binary, each partial product is either zero (multiplication by 0) or exactly same as the multiplicand (multiplication by 1). An example of binary multiplication is given below :

***Example 16.*** *Multiply 1001 by 1101.*

*Solution*

$$\begin{array}{rcl} 1\;0\;0\;1 & \mathrm{I} & \text{Multiplicand} \\ \times\quad 1\;1\;0\;1 & \mathrm{II} & \text{Multiplier} \\ \hline 1\;0\;0\;1 & \mathrm{III} & \\ 0\;0\;0\;0\phantom{\;0} & \mathrm{IV} & \text{Partial Products} \\ 1\;0\;0\;1\phantom{\;0\;0} & & \\ 1\;0\;0\;1\phantom{\;0\;0\;0} & & \\ \hline 1\;1\;1\;0\;1\;0\;1 & & \text{Final Product} \end{array}$$

In a digital circuit, the multiplication operation is performed by repeated additions of all partial products to obtain the full product.

**Binary Division**

Binary division is obtained using the same procedure as decimal division. An example of binary division is given below :

***Example 17 :*** Divide 1 1 1 0 1 0 1 by 1 0 0 1.

*Solution*

$$
\begin{array}{rl}
1101 & \leftarrow \text{Quotient} \\
\text{Divisor} \rightarrow 1001 \overline{)1110101} & \leftarrow \text{Dividend} \\
\underline{1001\phantom{000}} & \\
1011\phantom{00} & \\
\underline{1001\phantom{00}} & \\
001001 & \\
\underline{1001} & \\
0000 &
\end{array}
$$

Ans : 1101

## 2'S COMPLEMENT ARITHMETIC

Digital circuits are used for performing binary arithmetic operations. It is possible to use the circuits designed for binary addition to perform the binary subtraction also if we can change the problem of subtraction to that of an addition. This concept eliminates the need of additional circuits for subtraction, rather the same adder circuits are used for both the operations. This makes design of arithmetic circuits very convenient and cheaper. For this purpose, 2's complement representation is used.

### Subtraction Using 2's Complement

Binary subtraction can be performed by adding the 2's complement of the subtrahend to the minuend. If a final carry is generated, discard the carry and the answer is given by the remaining bits which is positive (the minuend is greater than the subtrahend). If the final carry is 0, the answer is negative (the minuend is smaller than the subtrahend ) and is in 2's complement form.

***Example 18 :*** Perform binary subtraction using 2's complement representation of negative numbers.

*Solution* (i)

$$
\begin{array}{rcrl}
7 & & 0111 & \text{Minuend} \\
\underline{-5} & \Rightarrow & \underline{(+)1011} & \text{2's complement of subtrahend} \\
+2 & & 10010 & \\
& & \uparrow & \\
& & \text{Discard final carry} &
\end{array}
$$

The answer is 0 0 1 0 equivalent $(+2)_{10}$.

(ii)

$$
\begin{array}{rcrl}
5 & & 0101 & \text{Minuend} \\
\underline{-7} & \Rightarrow & \underline{(+)1001} & \text{2's complement of subtrahend} \\
-2 & & 1110 &
\end{array}
$$

The final carry = 0. Therefore, the answer is negative and is in 2's complement form.
2's complement of 1 1 1 0 = 0 0 1 0

Therefore, the answer is $(-2)_{10}$

### Addition/Subtraction in 2's Complement Representation

The addition/subtraction of signed binary numbers can most conveniently be performed using 2's complement representation of both the operands. This is the method most commonly used when these operations are performed using digital circuits and microprocessors.

***Example 19 :*** Perform the following operations using 2's complement method :
(i) 48 - 23 (ii) 23 - 48 (iii) 48 - (-23) (iv) -48 - 23
Use 8-bit representation of numbers.
*Solution*
(i) 2's complement representation of + 48 = 0 0 1 1 0 0 0 0
2's complement representation of -23 = 1 1 1 0 1 0 0 1

48 + (-23)

$$
\begin{array}{rcrl}
 & 48 & & 00110000 \\
\Rightarrow & \underline{+\ (-23)} & \Rightarrow & \underline{(+)11101001} \\
 & +25 & & 100011001 \qquad = +25 \\
 & & & \uparrow \\
 & & & \text{Discard Carry}
\end{array}
$$

(ii) 2's complement representation of + 23 = 0 0 0 1 0 1 1 1
2's complement representation of - 48 = 1 1 0 1 0 0 0 0
23 - 48 = 23 + (-48)

$$
\begin{array}{rcrl}
 & 23 & & 00010111 \\
\Rightarrow & \underline{+\ (-48)} & \Rightarrow & \underline{(+)11010000} \\
 & -25 & & 11100111 \Rightarrow \qquad = -25
\end{array}
$$

(iii) 48 - (-23) = 48 + 23

$$
\begin{array}{rcrl}
 & 48 & & 00110000 \\
\Rightarrow & \underline{+\ (23)} & \Rightarrow & \underline{(+)\,00010111} \\
 & +71 & & 01000111 \Rightarrow \qquad = +71
\end{array}
$$

(iv) - 48 - 23 = (-48) + (-23)

$$
\begin{array}{rcrl}
 & -48 & & 11010000 \\
\Rightarrow & \underline{+\ (-23)} & \Rightarrow & \underline{(+)11101001} \\
 & -71 & & 110111001 \Rightarrow \qquad = -71 \\
 & & & \uparrow \\
 & & & \text{carry be ignored}
\end{array}
$$

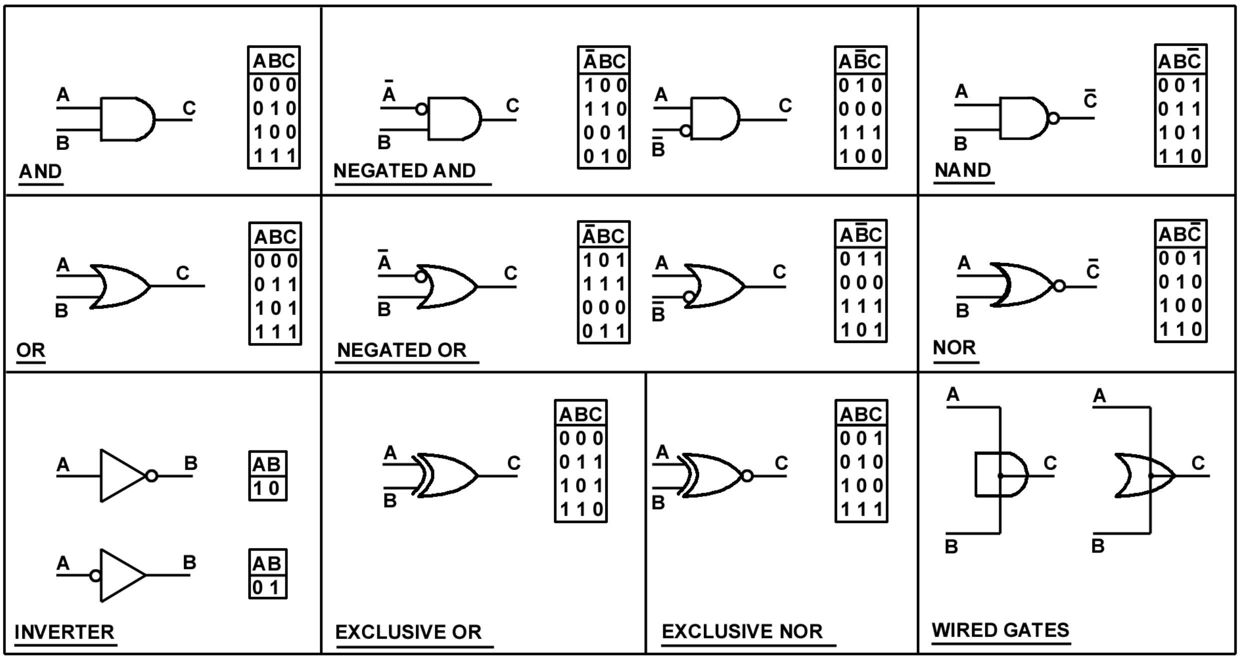
From the above example, we observe the following :

(a) If the two operands are of the opposite sign, the result is to be obtained by the rule of subtraction using 2's complement.

(b) If the two operands are of the same sign, the sign bit of the result (msb) is to be compared with the sign bit of the operands. In case the sign bits are same, the result is correct and is in 2's complement form. If the sign bits are not same there is a problem of over flow, i.e. the result can not be accommodated using eight bits and the result is to be interpreted suitably. The result in this case will consist of nine bits, i.e. carry and eight bits, and the carry bit will give the sign of the number.

■■■

# CHAPTER : 29
# LOGIC GATES AND TRUTH TABLES




■■■

# CHAPTER : 30
# ELEMENTARY KNOWLEDGE OF COMPUTERS, ITS APPLICATIONS

**A SIMPLE MODEL OF A COMPUTER.**
A computing machine designed to carry out algorithms for information processing has the configuration of Fig. 30.1. It is seen that an input unit is provided to read the algorithm and the data to be processed by the algorithm. The memory unit stores the algorithm and computed values. The processing unit interprets the instructions and carries them out. It has the capability to perform arithmetic operations, character manipulation operations, and logical operations. The output unit prints or displays computed results.

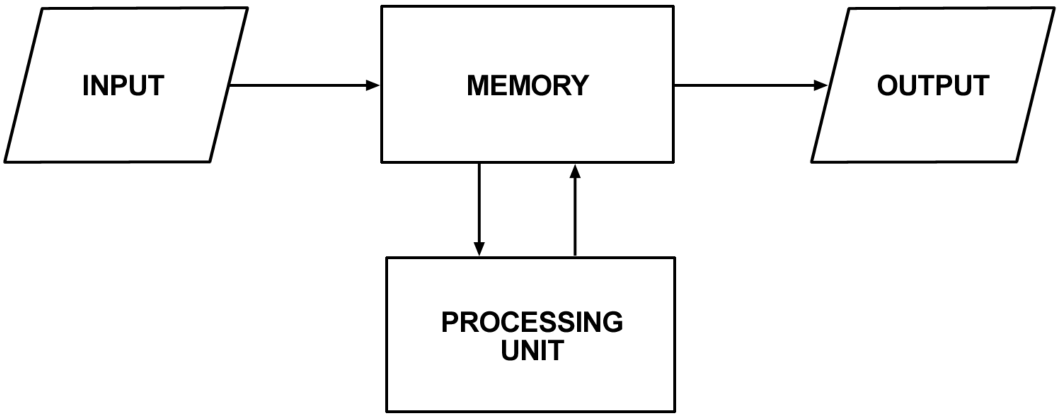



*Fig. 30.1, Organization of a computer.*

It was seen that it is possible to develop an enormous number of different interesting patterns by permuting and combining a few basic types of instructions. The same principle is used in building computers. Thus by using a computer's processing unit which can interpret and execute as few as ten different operations, it is possible to perform a large variety of information processing tasks.

**CHARACTERISTICS OF COMPUTERS**
The interesting features of a computer are :
1. Computers are built to carry out a small variety of instructions. It is not necessary to have more than about 100 distinct instructions even for a very powerful machine.
2. Instructions are extremely simple; e.g., add, subtract, read a character, write a character, compare numbers, characters, etc.
3. Most instructions are carried out in less than a millionth of a second.
4. Instructions are carried out obediently with no questions asked.
5. Instructions are carried out without any mistakes.

A computer may thus be thought of as a servant who would carry out instructions obediently, uncritically, at a very high speed, and without exhibiting any emotions. As human beings, we use judgement based on experience, often on subjective and emotional considerations. Such value judgements often depend on what is called sound "commonsense". As opposed to his, a computer exhibits no emotions and has no commonsense. An algorithm may be written for a computer to compose music based on rules of composition, but the computer cannot judge the quality of the resultant music. It must be clearly understood that computers are machines which can be programmed to follow instructions; they don't have their own priorities and judgements. Computers are machines which can help mankind in many ways; but they do not threaten us.

Being obedient without exercising ; commonsense; can be very annoying and unproductive. Take the instance of a Colonel who sent his obedient orderly to a post office with the order "Go to the Post Office and buy ten 25 paise stamps". The orderly went with the money to the post office and did not return for a long time. The Colonel got worried and went in search of him to the post office and found the orderly standing there with the stamps in his hand. When the colonel became angry and asked the orderly why he was standing there, pat came the reply that he was ordered to buy ten 25 paise stamps but not ordered to return with them !

A consequence of the uncritical acceptance of order by a computer is the need to give extensive, detailed, and correct instructions for solving problems. This can be quite challenging.

**PROBLEM-SOLVING USING COMPUTERS**

In order to solve a problem using a computer the following steps are followed :

1. The given problem is analysed.
2. The solution method is broken down into a sequence of elementary tasks.
3. Based on this analysis an algorithm to solve the problem is formulated. The algorithm should be precise, concise and unambiguous. Based on our discussions we realize that algorithm formulation is difficult and time-consuming.
4. The algorithm is expressed in a precise notation. An algorithm expressed using a precise notation is called a computer program. The precise notation is called a computer programming language.
5. The computer program is fed to the computer.
6. The computer's processing unit interprets the instructions in the program, executes them and sends the results to the output unit.

**FIRST GENERATION OF COMPUTERS**

The first electronic computer was completed in 1946 by a team led by Eckert and Mauchly at the University of Pennsylvania in U.S.A. This computer called Electronic Numerical Integrator and Calculator (ENIAC) used high speed vacuum tube switching devices. It had a very small memory and was designed primarily to calculate the trajectories of missiles. The ENIAC took about 200 microseconds to add two digits and about 2800 microseconds to multiply.

A major breakthrough occurred in the logical design of computers when the concept of a stored program was proposed by John Von Neumann in 1946. His idea was to store machine instructions in the memory of the computer along with data. These instructions could themselves be modified as required by other instructions. This allowed easy implementation of program loops. The first computer using this principle was designed and commissioned at Cambridge University, U.K. under the leadership of Professor Maurice Wilkes. This computer called EDSAC (Electronic Delay Storage Automatic Calculator) was completed in 1949 and used mercury delay lines for storage.

Commercial production of stored program electronic computers began in the early 50s. One of the early computers of this type was UNIVAC I built by Univac division of Remington Rand and delivered in 1951. This computer also used vacuum tubes. As vacuum tubes used filaments as a source of electrons, they had a limited life. Each tube consumed about half a watt power. Computers typically used about ten thousand tubes. Power dissipation was very high. As a large number of tubes, each with limited life, was used in fabricating these computers, their mean time between failures was low - of the order of an hour.

During this period, computer programming was mainly done in machine language. Assembly language was invented in the early fifties. Initial applications were in science and engineering. With the advent of UNIVAC, the prospects of commercial application were perceived. The concept of an operating system had not yet emerged. By and large during this period one had to be a good electronics engineer, and understand the logical structure of a computer in great detail, and also know how to program an application in order to use a computer. It was somewhat like the early days of motor cars when one had to be a good mechanic to be able to drive a car!

**THE SECOND GENERATION**

A big revolution in electronics took place with the invention of transistors by Bardeen, Brattain and Shockey in 1947. Transistors made of germanium semiconductor material were highly reliable compared to tubes since transistors had no filament to burn. They occupied less space and used only a tenth of the power required by tubes. They also could switch from a 0 to a 1 state in a few microseconds, about a tenth of the time needed by tubes. Thus switching circuits for computers made with transistors were about ten times more reliable, ten times faster, dissipated one tenth the power, occupied about one tenth the space and were ten times cheaper than those using tubes. Computer manufacturers thus changed over to transistors form tubes. The second generation computers emerged around 1955 with the use of transistors instead of vacuum tubes in computers. This generation lasted till 1965.

Another major event during this period was the invention of magnetic cores for storage. Magnetic cores are tiny rings (.002 inch diameter) made of ferrite and can be magnetized in either clockwise or anti-clockwise direction. The two directions are used to represent a 0 and a 1. Magnetic cores were used to construct large random access memories. Memory capacity in the second generation was about 100 kilo bytes. Magnetic disk storage was also developed during this period.

The higher reliability of computers and large memory availability led to the development of high level languages. Fortran, COBOL, Algol and Snobol were developed during this generation. With higher speed CPUs and the advent of magnetic tape and disk storage, operating systems were developed. Good batch operating systems, particularly the ones on IBM 7000 series computers emerged during the second generation.

Commercial applications rapidly developed during this period and dominated computer use by mid 1960s. More than 80% of installed computers were used in business and industry. All systems were batch oriented. Payroll, inventory control, marketing, production planning and general ledger systems were developed. A number of applications of operations research such as Linear Programming, Critical Path Methods (CPM) and Simulation became popular. Engineering applications, particularly in process control, increased rapidly.

New professions in computing such as systems analysts and programmers emerged during the second generation. Academic programmes in computer science were also initiated.

**THE THIRD GENERATION**

The third generation began in 1965 with germanium transistors being replaced by silicon transistors. Integrated circuits, circuits consisting of transistors, resistors, and capacitors grown on a single chip of silicon eliminating wired interconnection between components, emerged. From small scale integrated circuits which had about 10 transistors per chip, technology developed to medium scale integrated circuits with 100 transistors per chip. Switching speed of transistors went up by a factor of 10, reliability increased by a factor of 10, power dissipation reduced by a factor of 10 and size was also reduced by a factor of 10. The cumulative effect of this was the emergence of extremely powerful CPUs with the capacity of carrying out 1 million instructions per second.

There were significant improvements in the design of magnetic core memories. The size of main memories reached about 4 Megabytes. Magnetic disk technology improved rapidly. 100 Megabytes/drive became feasible.

The combined effect of high capacity memory, powerful CPU and large disk memories led to the development of time shared operating systems. Time shared systems increased programmer productivity.

Many important on-line systems became feasible. In particular dynamic production control systems, airline reservation systems, interactive query systems, and real-time closed loop process control systems were implemented. Integrated data base management systems emerged.

High level languages improved. Fortran IV and optimizing Fortran compilers were developed. COBOL 68 was standardized by the American National Standards Institute. PL/1 of IBM emerged and was quite a powerful language.

The third generation probably ended by 1975. The improvements in the period 65-75 were substantial but no revolutionary new concept could be identified as heralding the end of generation.

**THE FOURTH GENERATION**

**First Decade (1976-85)**

The fourth generation may be identified by the advent of the microprocessor chip. Medium scale integrated circuits yielded to Large and Very Large Scale Integrated circuits (VLSI) packing about 50000 transistors in a chip. Magnetic core memories were replaced by semiconductor memories. Semiconductor memory sizes of 16 Megabytes with a cycle time of 200 nsecs were in common use. The emergence of the microprocessor led to two directions in computer development. One direction was the emergence of extremely powerful personal computers. Computer cost came down so rapidly that professionals had their own computer to use in their office and Hard disks provided a low cost, high capacity secondary memory.

The other direction of development was the decentralization of computer organization. Individual microprocessor controls for terminals and peripheral devices allowed the CPU to concentrate on processing the main program. Networks of computers and distributed computer systems were developed. Disk memories became very large (1000 Mbytes/drive). A significant development in software was the development of concurrent programming languages. Such languages are important to program distributed systems and real time systems. The most ambitious language of this type was ADA. Another important development was interactive graphic devices and language interfaces to graphic systems. The emergence of graphics gave a great impetus to computer-aided engineering design.

Fourth generation saw the coming of age of UNIX OS and time shared interactive systems. These systems became user friendly and highly reliable. The effective cost of computing came down. Computers also became all pervading.

**Second Phase (1986-2000)**

The second phase of the fourth generation has seen a relentless increase in the speed of microprocessors and the size of main memory. The speed of microprocessors and the size of main memory and hard disk went up by a factor of 4 every 3 years. Many of the features originally found in CPUs of large expensive mainframe computers of the first decade of the fourth generation became part of the microprocessor architecture in the 90s. Thus the mainframe computer of early 80s died in mid 90s. The alpha microprocessor chip designed by DEC in 1994 packed 9.3 million transistors in a single chip, was driven by a 300 MHz clock and could carry out a billion operations per second. It had a built-in 64-bit floating point arithmetic unit, used 64-bit data and 64-bit address buses. It had a built-in cache memory of 64 KB and 32 registers to store temporary operands. Apart from this IBM, Apple computers and Motorola cooperated in designing a microprocessor called Power PC 600 series. Intel also designed a powerful chip in 90s called Pentium (1993) which sold in large numbers. The original Pentium was followed by Pentium with MMX (Multimedia Extension) and Pentium II with a clock speed of 466 MHz and Celeron processor with a 300 MHz clock. In 2000 Intel introduced a 64-bit processor called IA 64 or Itanium.

Microprocessors such as Pentium, Power PC, etc., are being used as the CPU of Personal Computers and portable laptop and palm held computers. Desk top workstations and powerful servers for numeric computing as well as file services use RISC microprocessors such as Alpha, MIPS and SUNSPARC.

The area of hard disk storage also saw vast improvements. 1 GB of disk on workstations became common in 1994. For larger disks RAID technology (Redundant Array of Inexpensive Disks) was used to give storage of 100 GB. Optical disks also emerged as mass storage particularly for read only files.

Optical storage sizes were of the order of 600 MB on a 5.25" disk. New optical disks known as Digital Versatile Disk ROMs, (DVD ROMs) with maximum storage capacity of around 17 GB emerged around 1998. Writable CDs were developed around the same time. The availability of optical disks at low cost saw the development of multimedia applications. Multimedia workstations were widely used.

Computer Networks came of age. The networks became very powerful with the advent of fibre optic Local Area Networks which could transmit 100 MB/sec to 1 GB/sec. Many mainframes were replaced by powerful workstations connected by fibre optic network. Another major event during this phase was the rapid increase in the number of computers connected to the internet. This led to the emergence of the World Wide Web which eased information retrieval. The Internet also brought out the need to execute programs on a variety of computers. This led to the emergence of a new object oriented language Java. Applications written in Java, called Java applets, could be glued together with a software called Java script to create large programs.

In the area of languages C language became popular. This was followed by a new method of design called object oriented design. The primary objectives of object oriented design are to generalize programs and to reuse objects. The C++ language emerged as the most popular object oriented language. One also saw a trend towards design of specification oriented languages. PROLOG was designed for logic oriented specification language and HASKELL, FP etc., as functional specification oriented language. With the emergence of distributed computers connected by networks considerable effort has gone into programming distributed systems. A number of parallel computers were built but no commonly accepted standard parallel programming language emerged.

**THE FIFTH GENERATION**

It is not very clear now what direction the fifth generation will take. It is estimated that by 2005 we may see computers of this generation.

Even though computers in the last 50 years have become very fast, reliable and inexpensive, the basic logical structure proposed by Von Neumann has not changed. The basic block diagram of a CPU, memory and I/O is still valid today. With the improvements in integrated circuit technology, it is now possible to get specialized VLSI chips at a low cost. Thus an architecture which makes use of the changes in technology and allows an easier and more natural problem solving is being sought. In Table 30.1 we summarize and compare various generation of computers.

**Table 30.1, Computer Generations - A comparison**

| Generation | Years | Switching device | Storage device | Switching time | MTBF* | Software | Applications |
|---|---|---|---|---|---|---|---|
| First | 1949-55 | Vacuum tubes | Acoustic delay lines and later magnetic drum. 1 Kbyte memory | 0.1 to 1 milli-second | 30 mts. to 1 hour | Machine and assembly languages. simple monitors | Mostly scientific. Later simple business systems |
| Second | 1956-65 | Transistors | Magnetic core main memory, tapes and disk peripheral memory. 100 Kbyte main memory | 1 to 10 micro-seconds | About 10 hrs. | High level languages. FORTRAN, COBOL, Algol, Batch Operating systems | Extensive business applications. Engineering design optimization, scientific research |
| Third | 1966-75 | Integrated Circuits (IC) | High speed magnetic cores. Large disks (100 MB). 1 1 Mbyte main memory | 0.1 to 1 micro-second | About 100 hrs. | FORTRAN IV, COBOL 68, PL/1. Timeshared operating system | Data base management systems. On-line systems |
| Fourth - First phase | 1975-84 | Large scale integrated circuits. Microprocessors (LSI) | Semiconductor memory. winches - ter disk. 10 Mbyte main memory. 1000 Mbyte disks | 10 to 100 nano-seconds | About 1000 hrs. | FORTRAN 77, Pascal, ADA, COBOL-74, Concurrent Pascal | Personal computer. Distributed systems. Integrated CAD/ CAM Real time control. Graphics oriented systems |
| Fourth - Second phase | 1985-present | Very large scale integrated circuits. Over 100 million transistors per chip | Semiconductor memory. 1GB main memory. 100 GB disk. | 1 to 10 nano-seconds | About 10,000 hrs. | C, C++, JAVA, PROLOG, Haskell Fortran 90/95 | Simulation, Visualization, Parallel computing, Virtual reality, Multimedia |

* MTBF - Mean time between failures of the processor.

## CLASSIFICATION OF COMPUTERS

Until recently computers were classified as microcomputers, minicomputers, supermini computers, mainframes, and supercomputers. Technology, however, has changed and this classification is no more relevant. Today all computers use microprocessors as their CPU. Thus classification is possible only through their mode of use. Based on mode of use we can classify computers as Palmtops, Laptop PCs, Desktop PCs and Workstations. Based on interconnected computers we can classify them as distributed computers and parallel computers.

### Palmtop PCs

With miniaturization and high density packing of transistors on a chip, computers with capabilities nearly that of PCs which can be held in a palm have emerged. Palmtops accept handwritten inputs using an electronic pen which can be used to write on a Palmtop's screen (besides a tiny keyboard), have small disk storage and can be connected to a wireless network. One has to train the system on the user's handwriting before it can be used. A Palmtop computer has also facilities to be used as a mobile phone, Fax and email machine. A version of Microsoft operating system called Windows is available for Palmtops.

### Laptop PCs

Laptop PCs (also known as notebook computers) are portable computers weighing around 2 kg. They use a keyboard, flat screen liquid crystal display, and a Pentium or Power PC processor. Colour displays are available. They normally run WINDOWS OS. Laptops come with both hard disk (around 1 GB), CDROM and floppy disk. They should run with batteries and are thus designed to conserve energy. Many Laptops can be connected to a network. There is a trend towards providing wireless connectivity to Laptops so that they can read files form large stationary computers. The most common use of Laptop computers is for word processing, and spreadsheet computing while a person is travelling. As Laptops use miniature components which have to consume low power and have to be packaged in small volume they cost 3 to 4 times the cost of table top PCs of the same capacity.

### Personal Computers (PCs)

The most popular PCs are desktop machines. Early PCs had intel 8088 microprocessors as their CPU. Currently (2001), intel Pentium IV is the most popular processor. The machines made by IBM are called IBM PCs. Other manufacturers use IBM's specifications and design their own PCs. They are known as IBM compatible PCs. IBM PCs mostly use MS DOS or MS-Windows, WINDOWS - NT or UNIX as Operating System. An OS called OS/2 is available for IBM PCs, and is also widely used. IBM PCs, nowadays (2001) have 64 to 256 MB main memory, 10 to 20 GB of disk and a floppy disk. Besides these a 600 MB optical disk is also provided in PCs intended for multimedia use. PCs are also made by another company called Apple. Apple PCs are known as Apple Macintosh. They use Apple's proprietary OS which is designed for simplicity of use. Apple Macintosh machines used Motorola 68030 microprocessors but now use Power PC 603 processor. IBM PCs are today the most popular computers with millions of them in use throughout the world.

### Workstations

Workstations are also desktop machines. They are, however, more powerful providing processor speeds about 10 times that of PCs. Most workstations have a large colour video display unit (19 inch monitors). Normally they have main memory of around 256 MB to 1 GB and disk of 20 to 40 GB. Workstations normally use RISC processors such as MIPS (SIG), ALPHA (DEC), RIOS (IBM), SPARC (SUN) OR PA - RISC (HP). Some manufacturers of Workstations are Silicon Graphics (SIG), Digital Equipment Corporation (DEC), IBM, SUN Microsystems and Hewlett Packard (HP). The standard Operating System of Workstations is UNIX and its derivatives such as AIX (IBM), Solaris (SUN), and HP-UX (HP). Very good graphics facilities are provided by most workstations. A system called X Windows is provided by Workstations to display the status of multiple processes during their execution. Most workstations have built-in hardware to connect to a Local Area Network (LAN). Workstations are used for executing numeric and graphic intensive applications such as those which arise in Computer aided Design, simulation of complex systems and visualizing the results of simulation.

### Mainframe Computers

There are organizations such as banks and insurance companies which process large number of transactions on-line. They require computers with very large disks to store several Tera bytes of data and transfer data from disk to main memory at several hundred Megabytes/sec. The processing power needed from such computers is hundred million transactions per second. These computers are much bigger and faster than workstations and several hundred times more expensive. They normally use proprietary operating systems which usually provide extensive services such as user accounting, file security and control. They are normally much more reliable when compared to operating systems on PCs. These types of computers are called mainframes. There are a few manufacturers of mainframes (e.g. IBM and Hitachi). The number of mainframe users has reduced as many organizations are rewriting their systems to use networks of powerful workstations.

### Supercomputers

Supercomputers are the fastest computers available at any given time and are normally used to solve problem which require intensive numerical computations. Examples of such problems are numerical weather prediction, designing supersonic aircrafts, design of drugs and modelling complex molecules. All of these problems require around $10^5$ calculations to be performed. Such a problem will be solved in about 3 hours by a computer which can carry out 100 billion floating point calculations per second. Such a computer is classified as a supercomputer today (2001). By about the year 2005 computers which can carry out $10^{15}$ floating point operations per second on 64 bit floating point numbers would be available and would be the ones which will be called supercomputers. Such a computer is built by interconnecting several high speed computers and programming them to work co-operatively to solve problems. Recently applications of supercomputers have expanded beyond scientific computing. They are now used to analyse large commercial databases, produce animated movies and play games such as chess.

Besides arithmetic speed, a computer to be classified as a supercomputer, should have large main memory of around 16 GB and a secondary memory of 1000 GB. The speed of transfer of information from the secondary memory to the main memory should be at least a tenth of the memory to CPU data transfer speed. All supercomputers use parallelism to achieve their speed.

■■■

# CHAPTER : 31
# SPECTRUM OF WAVES

**FREQUENCY SPECTRUM OF THE AM WAVE**

We shall show mathematically that the frequencies present in the AM wave are the carrier frequency and the first pair of sideband frequencies, where a sideband frequency is defined as

$$f_{SB} = f_c \pm nf_m \quad (1)$$

and in the first pair n = 1

When a carrier is amplitude-modulated, the proportionality constant is made equal to unity, and the instantaneous modulating voltage variations are superimposed onto the carrier amplitude. Thus, when there is temporarily no modulation, the amplitude of the carrier is equal to its unmodulated value. When modulation is present, the amplitude of the carrier is varied by its instantaneous value. The situation is illustrated in Fig.31.1(a), which shows how the maximum amplitude of the amplitude-modulated voltage is made to vary in accordance with modulating voltage change. Fig.31.1(a), also show that something unusual (distortion, as it happens) will occur if $V_m$ is greater than $V_c$. This, and the fact that the ratio $\frac{V_m}{V_c}$ often occurs, leads to the following definition of the modulation index:

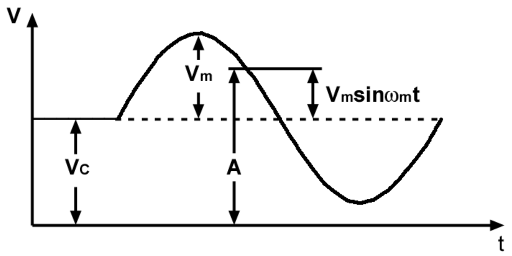



*Fig. 31.1 (a), Amplitude of AM wave.*

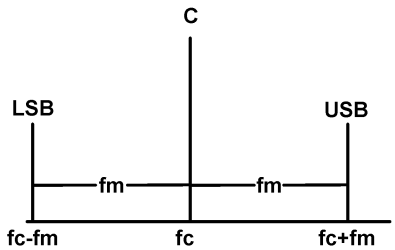



*Fig. 31.1 (b), Frequency Spectrum of AM wave.*

$$m = \frac{V_m}{V_c} \quad (2)$$

The modulation index is a number lying between 0 and 1, and it is very often expressed as a percentage and called the percentage modulation.

From Fig. 31.1(a) and Eq. 2 it is possible to write an equation for the amplitude of the amplitude-modulated voltage. Thus we have

$$A = V_c + v_m = V_c + V_m \sin \omega_m t = V_c + mV_c \sin \omega_m t$$
$$= V_c ( 1 + m \sin \omega_m t ) \quad (3)$$

The instantaneous voltage of the resulting amplitude-modulated wave is

$$v = A \sin \theta = A \sin \omega_c t = V_c ( 1+ m \sin \omega_m t ) \sin \omega_c t \quad (4)$$

Equation 4 may be expanded, by means of the trigonometrical relation

$\sin x \sin y = ½ [ \cos (x-y) - \cos (x + y)]$, to give

$$v = V_c \sin \omega_c t + \frac{mV_c}{2} \cos(\omega_c + \omega_m)t - \frac{mV_c}{2} \cos(\omega_c - \omega_m)t \quad (5)$$

It has thus been shown that the equation of an amplitude-modulated wave contains three terms. The first term represents the unmodulated carrier. It is thus apparent that the process of amplitude modulation has the effect of adding to the unmodulated wave, rather than changing it. The two additional terms produced are two sidebands outlined. The frequency of the lower sideband (LSB) is $f_c - f_m$, and the frequency of the upper sideband (USB) is $f_c + f_m$. The very important conclusion to be made at this stage is that the bandwidth required for amplitude modulation is twice the frequency of the modulating signal. In modulation by several sine waves simultaneously, as in the AM broadcasting service, the bandwidth required is twice the highest modulating frequency [Fig. 31.1 (b)]

**FREQUENCY SPECTRUM OF THE FM WAVE**

When a comparable stage was reached with AM theory, i.e., when Eq. (5) has been derived, it was possible to tell at a glance what frequencies were present in the modulated wave. Unfortunately, the situation is far more complex, mathematically speaking, for FM. Since Eq. $V = A \sin (\omega_c t + m_f \sin\omega_m t)$ is the sine of a sine, the only solution involves the use of Bessel functions. Using these, it may then be shown that Eq. above may be expanded to yield

$$v = A\{J_o(m_f)\sin\omega_c t$$
$$+ J_1(m_f)[\sin(\omega_c + \omega_m)t - \sin(\omega_c - \omega_m)t]$$
$$+ J_2(m_f)[\sin(\omega_c + 2\omega_m)t + \sin(\omega_c - 2\omega_m)t]$$
$$+ J_3(m_f)[\sin(\omega_c + 3\omega_m)t - \sin(\omega_c - 3\omega_m)t]$$
$$+ J_4(m_f)[\sin(\omega_c + 4\omega_m)t + \sin(\omega_c - 4\omega_m)t] \ldots\} \quad (1)$$

It is seen that the output consists of a carrier and an apparently infinite number of pairs of sidebands, each preceded by J coefficients. These are Bessel function. Here they happen to be of the first kind and of the order denoted by the subscript, with the argument $m_f$. $J_n(m_f)$may be shown to be a solution of an equation of the form

$$(m_f)^2 (d^2y / dmf^2) + m_f (dy/dm_f) + (m_f^2 - n^2)y = 0 \quad (2)$$

This solution, i.e., the formula for the Bessel function, is

$$j_n(m_f) = (m_f/2)^n \,[\,(1/n!) - (m_f/2)^2/1!\,(n+1)! + (m_f/2)^4/2!(n+2)! - (m_f/2)^6 / 3!(n+1)! + \ldots\} \quad (3)$$

In order to evaluate the value of a given pair of sidebands or, the value of the carrier, it is necessary to know the value of the corresponding Bessel function. However, separate calculation from Eq.3 for each case is not required since information of this type is freely available in table form, as in Table 31.1, or graphical form, as in Fig. 31.2.

**Observations**

The mathematics of the foregoing discussion may be reviewed in a series of observations as follows:

1. Unlike AM, where there are only three frequencies (the carrier and the first two sidebands), FM has an infinite number of sidebands, as well as the carrier. They are separated from the carrier by $f_m$, $2f_m$, $3f_m$, .... , and thus have a recurrence frequency of $f_m$.

**TABLE 31.1, BESSEL FUNCTIONS OF THE FIRST KIND**

| x (mf) | n or Order | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $J_0$ | $J_1$ | $J_2$ | $J_3$ | $J_4$ | $J_5$ | $J_6$ | $J_7$ | $J_8$ | $J_9$ | $J_{10}$ | $J_{11}$ | $J_{12}$ | $J_{13}$ | $J_{14}$ | $J_{15}$ | $J_{16}$ |
| 0.00 | 1.00 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 0.25 | 0.98 | 0.12 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 0.5 | 0.94 | 0.24 | 0.03 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 1.0 | 0.77 | 0.44 | 0.11 | 0.02 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 1.5 | 0.51 | 0.56 | 0.23 | 0.06 | 0.01 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 2.0 | 0.22 | 0.58 | 0.35 | 0.13 | 0.03 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 2.5 | -0.05 | 0.50 | 0.45 | 0.22 | 0.07 | 0.02 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 3.0 | -0.26 | 0.34 | 0.49 | 0.31 | 0.13 | 0.04 | 0.01 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 4.0 | -0.40 | -0.07 | 0.36 | 0.43 | 0.28 | 0.13 | 0.05 | 0.02 | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 5.0 | -0.18 | -0.33 | 0.05 | 0.36 | 0.39 | 0.26 | 0.13 | 0.05 | 0.02 | -- | -- | -- | -- | -- | -- | -- | -- |
| 6.0 | 0.15 | -0.28 | -0.24 | 0.11 | 0.36 | 0.36 | 0.25 | 0.13 | 0.06 | 0.02 | -- | -- | -- | -- | -- | -- | -- |
| 7.0 | 0.30 | 0.00 | -0.30 | -0.17 | 0.16 | 0.35 | 0.34 | 0.23 | 0.13 | 0.06 | 0.02 | -- | -- | -- | -- | -- | -- |
| 8.0 | 0.17 | 0.23 | -0.11 | -0.29 | -0.10 | 0.19 | 0.34 | 0.32 | 0.22 | 0.13 | 0.06 | 0.03 | -- | -- | -- | -- | -- |
| 9.0 | -0.09 | 0.24 | 0.14 | -0.18 | -0.27 | -0.06 | 0.20 | 0.33 | 0.30 | 0.21 | 0.12 | 0.06 | 0.03 | 0.01 | -- | -- | -- |
| 10.0 | -0.25 | 0.04 | 0.25 | 0.06 | -0.22 | -0.23 | -0.01 | 0.22 | 0.31 | 0.29 | 0.20 | 0.12 | 0.06 | 0.03 | 0.01 | -- | -- |
| 12.0 | 0.05 | -0.22 | -0.08 | 0.20 | 0.18 | -0.07 | -0.24 | -0.17 | 0.05 | 0.23 | 0.30 | 0.27 | 0.20 | 0.12 | 0.07 | 0.03 | 0.01 |
| 15.0 | -0.01 | 0.21 | 0.04 | -0.19 | -0.12 | 0.13 | 0.21 | 0.03 | -0.17 | -0.22 | -0.09 | 0.10 | 0.24 | 0.28 | 0.25 | 0.18 | 0.12 |

2. The J coefficients eventually decrease in value as n increases, but not in any simple manner. As seen in Fig. 31.2, the value fluctuates on either side of zero, gradually diminishing. Since each J coefficient represents the amplitude of a particular pair of sidebands, these also eventually decrease, but only past a certain value of n. The modulation index thus determines how many sideband components have significant amplitudes.
3. The sidebands at equal distances from $f_c$ have equal amplitudes, so that the sideband distribution is symmetrical about the carrier frequency. The J coefficients occasionally have negative values, signifying a 180° phase change for that particular pair of sidebands.
4. Looking down Table , we see that, as mf increases, so does the value of a particular J coefficient, such as (say) $J_{12}$. Bearing in mind that $m_f$ is inversely proportional to the modulating frequency, we see that the relative amplitude of distant sidebands increases when the modulation frequency is lowered. The fore going assumes that deviation (i.e., the modulating voltage ) has remained constant.
5. In AM, increased depth of modulation increases the sideband power always remains the total transmitted power. In FM, the total transmitted power always remains constant, but with increased depth of modulation the required bandwidth is increased. To be quite specific, what increases is the bandwidth required to transmit a relatively undistorted signal. This is true because increased depth of modulation means increased deviation, and therefore an increased modulation index, so that more distant sidebands acquire significant amplitudes.
6. As evidenced by Eq. (1), the theoretical bandwidth required in FM is infinite. In practice, the bandwidth used is one that has been calculated to allow for all significant amplitudes of sideband components under the most exacting conditions. This really means ensuring that, with maximum deviation by the highest modulating frequency, no significant sideband components are lopped off.
7. In FM, unlike in AM, the amplitude of the carrier component does not remain constant. Its J coefficient is $J_o$, which is, of course, a function of $m_f$. In fact, it is quite logical and necessary that this should be so. Since the overall amplitude to the FM wave remains constant, it would be very odd indeed if the amplitude of the carrier were not reduced when the amplitude of the various sidebands is increased, and vice versa.
8. It is possible for the carrier component of the FM wave to disappear completely. This happens for certain values

of the modulation index, called eigenvalues. Fig. 31.2, shows that these are approximately 2.4, 5.5, 8.6, 11.8, and so on. These disappearances of the carrier for specific values of $m_f$ form a handy basis for measuring deviation, as will be seen.

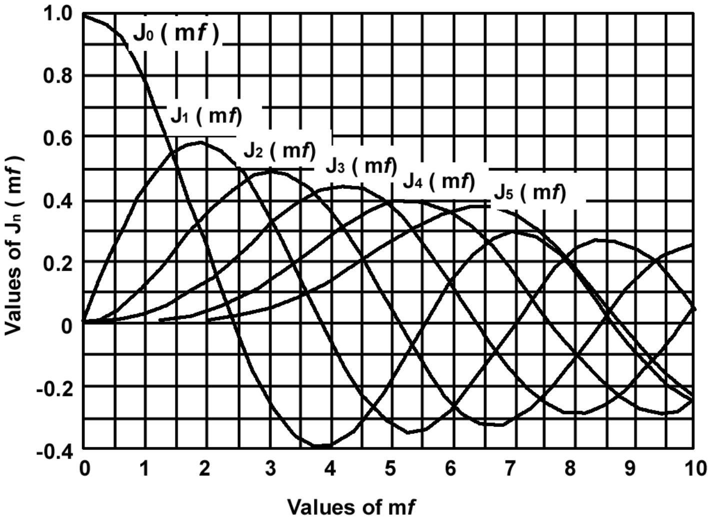



*Fig. 31.2, Bessel functions.*

**Bandwidth and required spectra**

Using Table 31.1, it is possible to evaluate the size of the carrier and each sideband for each specific or interesting value of the modulation index. When this is done, the frequency spectrum of the FM wave for that particular value of $m_f$ may be plotted. This is done in Fig. 31.3, which shows these spectrograms first for increasing deviation ($f_m$-constant), and then for decreasing modulating frequency (δ constant). Both the table and the spectrograms illustrate the observations, especially points 2, 3, 4, and 5. It is seen, for example, that as modulation depth increases, so does bandwidth (Fig. 31.3), and also that reduction in modulation frequency increases the number of sidebands, though not necessarily the bandwidth. Another point shown very clearly is that although the number of sideband components is theoretically infinite, in practice a lot of the higher sidebands have insignificant relative amplitudes, and this is why they are not shown in the spectrograms. Thus their exclusion in practical system will not distort the modulated wave unduly.

In order to calculate the required bandwidth accurately, one need merely glance at the table to see which is the last J coefficient shown for that value of modulation index.

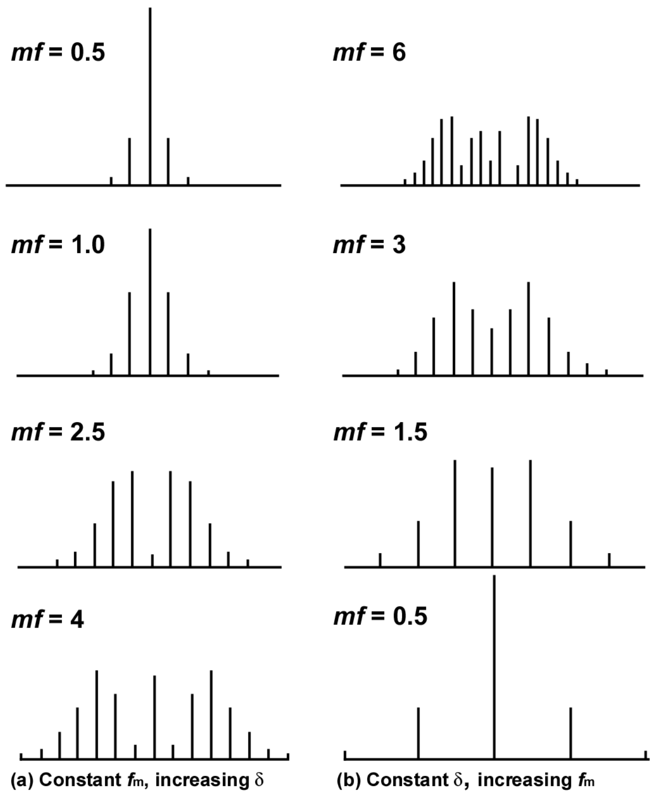



*Fig. 31.3, FM spectrograms. (After K.R. Sturley, Frequency-Modulated Radio, 2d ed., George Newnes Ltd., London, 1958, by permission of the publisher.)*

A rule of thumb (Carson's rule) states that (as a good approximation) the band width required to pass an FM wave is twice the sum of the deviation and the highest modulating frequency, but it must be remembered that this is only an approximation. Actually, it does give a fairly accurate result if the modulation index in excess of about 6.

**STEREOPHONIC FM MULTIPLEX SYSTEM**

Stereo FM transmission is a modulation system in which sufficient information is sent to the receiver to enable it to reproduce original stereo material. It became commercially available in 1961, several years after commercial monaural transmissions. Like colour TV (which of course came after monochrome TV), it suffers from the disadvantage of having been made more complicated than it needed to be, to ensure that it would be compatible with the existing system. Thus, in stereo FM, it is not possible to have a two-channel system a left channel and a right channel transmitted simultaneously and independently, because a monaural system would not receive all the information in an acceptable form.

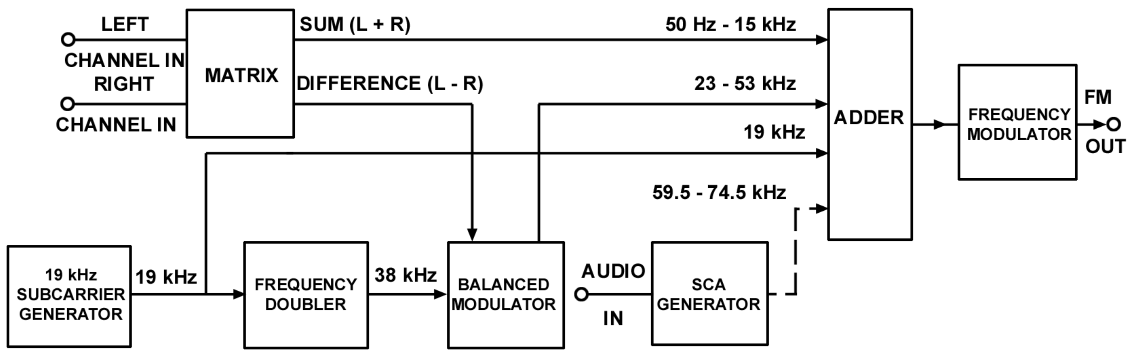



*Fig. 31.4, Stereo FM multiplex generator with optional SCA.*

As shown in the block diagram of Fig. 31. 4, the two channels in the FM stereo multiplex system are passed through a matrix which produces two outputs. The sum (L+R) modulates the carrier in the same manner as the signal in a monaural transmission, and this is the signal which is demodulated and reproduced by a mono receiver, tuned to a stereo transmission. The other output of the matrix is the difference signal (L-R). After demodulation in a stereo receiver, (L-R) will be added to (L+R) to produce the left channel, while the difference between the two signals will produce the right channel. It is necessary to understand how the difference signal is impressed on the carrier.

What happens, in essence, is that the difference signal is shifted in frequency from the 50-Hz to 15, 000-Hz range (which it would otherwise co-occupy with the sum signal) to a higher frequency. As will be seen, such signal " stacking" is known as multiplexing, hence the name of the system. In this case, as in other multiplexing, a form of single sideband suppressed carrier (SSBSC) is used, with the signals to be multiplexed up being modulated onto a sub carrier at a high audio or supersonic frequency. However, there is a snag here, which makes this form of multiplexing different from the more common ones. The problem is that the lowest audio frequency is 50 Hz, much lower than the normal minimum of 300 Hz encountered in communications voice channels. This makes it difficult to suppress the unwanted sideband without affecting the wanted one; pilot carrier extraction in the receiver is equally difficult. And yet some form of carrier must be transmitted, to ensure that the receiver has a stable reference frequency for demodulation: otherwise, distortion of the difference signal will result..

The two problems are solved in associated but separated ways. In the first place, the difference signal is applied to a balanced modulator (as it would be in any multiplexing system) which, in orthodox fashion, suppresses the carrier. However, both sidebands are then used as modulating signals and duly transmitted, whereas normally one might expect one of them to be removed prior to transmission. Since the sub carrier frequency is 38 kHz, the sidebands produced by the different signal occupy the frequency range from 23 to 53 kHz. It is thus seen that they do not interfere with the sum signal, which occupies the range of 50 Hz to 15 kHz.

The reason that the 38-kHz sub carrier is generated by a 19-kHz oscillator whose frequency is then doubled may now be explained. Indeed, this is the trick used to avoid the difficulty to having to extract the pilot carrier from among the close sideband frequencies in the receiver. As shown in the block diagram, the output of the 19-kHz sub carrier generator is added to the sum and difference signals in the output adder preceding the modulator. In the receiver, the frequency of the 19-kHz signal is doubled, and it can then be re-inserted as the carrier for the defference signal. It should be noted that the sub carrier is inserted at a level of 10 percent, which is both adequate and not so large as to take undue power from the sum and difference signals (or to cause over modulation). Also, the frequency of 19 kHz fits neatly into the space between the top of the sum signal and the bottom of the difference signal-it is far enough from each of them so that no difficulty in extracting it is experienced in the receiver..

The FM stereo multiplex system described here is the one used in the Untied States, and is in accordance with the standards established by the Federal Communications Commission (FCC) in 1961. Stereo FM has by now spread to broadcasting in most other parts of the world, where the systems in use are either identical or quite similar to the above. A subsidiary Communications Authorization (SCA) signal may also be transmitted in the U.S. stereo multiplex system; it is the remaining signal feeding in to the output adder. It is shown dashed in the diagram because it is not always present. Some stations provide SCA as a second, medium-quality transmission, used as background music in stores, restaurants and the like.

SCA uses a sub carrier at 67 kHz, modulated to a depth of $\pm$ 7.5 kHz by the audio signal. Frequency modulation is used. The frequency band thus occupied ranges from 59.5 to 74.5 kHz and fits sufficiently above the difference signal as not to interfere with it. The overall frequency allocation within the modulating signal of an FM stereo multiplex transmission with SCA is shown in Fig. 31.5. As can be appreciated, the amplitude of the sum and difference signals must be reduced (generally by 10 percent) in the presence of SCA; otherwise, over modulation of the main carrier could result.

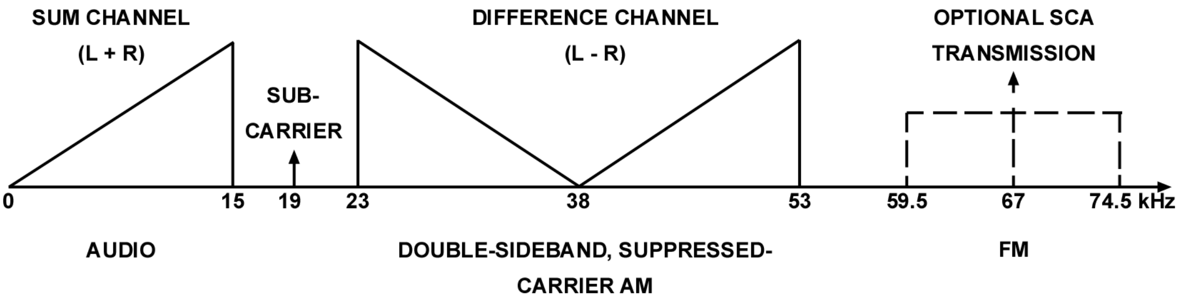



*Fig. 31.5, Spectrum of stereo FM multiplex modulating signal (with optional SCA).*

## PROPAGATION OF WAVES

In an earth environment, electromagnetic waves propagate in ways that depend not only on their own properties but also on those of the environment itself; some of this was seen in the preceding section. Since the various methods of propagation depend largely on frequency, the complete electromagnetic spectrum is now shown for reference in Fig.31.6 -note that the frequency scale is logarithmic.

Wave travel in straight lines, except where the earth and its atmosphere alter their path. Thus, except in unusual circumstances, frequencies above the HF generally travel in straight lines. They propagate by means of so-called space waves. These are sometimes called tropospheric waves, since they travel in the troposphere, the portion of the atmosphere closest to the ground. Frequencies below the HF range travel around the curvature of the earth, sometimes right around the globe. The means are probably a combination of diffraction and a type of wave guide effect which uses the earth's surface and the lowest ionized layer of the atmosphere as the two wave guide walls. These ground waves, or surface waves as they are called, are one of the two original-means of beyond-the-horizon propagation. For example, all broadcast radio signals received in daytime propagate by means of surface waves.

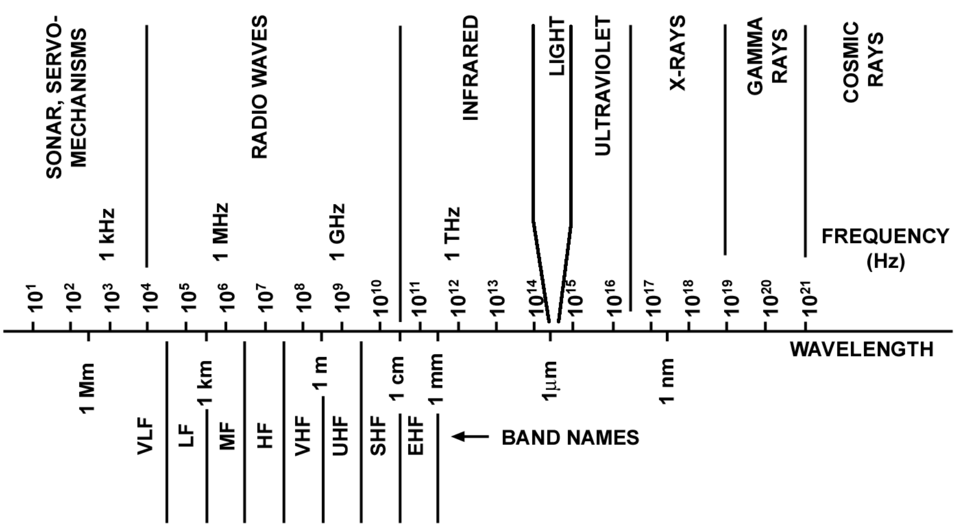



*Fig. 31.6, The electromagnetic spectrum.*

Wave in the HF range, and sometimes frequencies just above or below it, are reflected by the ionized layers of the atmosphere (to be described) and are called sky waves. Such signals are beamed into the sky and come down again after reflection, returning to earth well beyond the horizon. To reach receivers on the opposite side of the earth, these waves must be reflected by the ground and the ionosphere several times. It should be mentioned that neither surface waves nor sky waves are possible in space or on airless bodies such as the moon.

Two more recently developed means of beyond-the-horizon propagation are tropospheric scatter and stationary satellite communications. Each of these five methods of propagation will now be described in turn.

**Ground (Surface) Waves**

Ground waves progress along the surface of the earth and, as previously mentioned, must be vertically polarized to prevent short circuiting the electric component. A wave induces currents in the ground over which it passes and thus loses some energy by absorption. However, this is made up, to a certain extent, by energy diffracted downward from the upper portions of the wavefront.

There is another way in which the surface wave is attenuated: because of diffraction, the wavefront gradually tilts over, as shown in Fig. 31.7, As the wave propagates over the earth, it tilts over more and more, and the increasing tilt causes greater short circuiting of the electric field component of the wave and hence field strength reduction. Eventually, at some distance (in wavelengths) from the antenna, as partly determined by the type of surface over which the ground wave propagates, the wave "lies down and dies". It is important to realize this, since it shows that the maximum range of such a transmitter depends on its frequency as well as its power. Thus, in the VLF band, insufficient range of transmission can be cured by increasing the transmitting power. On the other hand, this remedy will not work near the top of the MF range, since propagation is now definitely limited by tilt.

**Field strength at a distance**

Radiation from an antenna by means of the ground wave gives rise to a field strength at a distance, which may be calculated by use of Maxwell's equations. This field strength, in volts per meter, is given in Eq.1, which differs from

$$\varepsilon = \frac{\sqrt{30 P_t}}{r} \tag{1}$$

by taking into account the gain of the transmitting antenna.

$$\varepsilon = \frac{120\pi h_t I}{\lambda d} \tag{2}$$

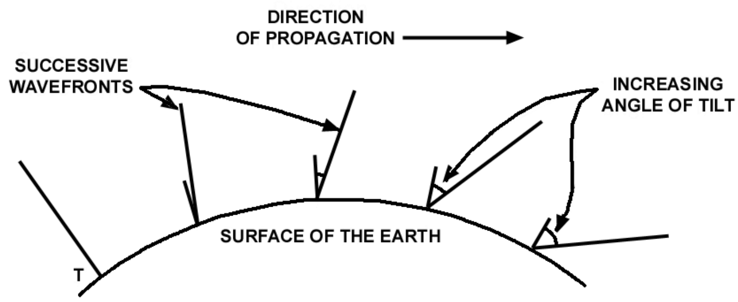



*Fig. 31.7, Ground-wave propagation.*

If a receiving antenna is now placed at this point, the signal it will receive will be, in volts,

$$V = \frac{120\pi h_t h_r I}{\lambda d} \tag{3}$$

where $120\pi$ = characteristic impedance of free space

$h_t$ = effective height of the transmitting antenna (this is no quite the same as actual height.)

$h_r$ = effective height of the receiving antenna

$I$ = antenna current

$d$ = distance from transmitting antenna

$\lambda$ = wavelength

If the distance between the two antennas is fairly long, the reduction of field strength due to ground and atmospheric absorption reduces the value of the voltage received, making it less than shown by Eq. (3). Although it is possible to calculate the signal strength reduction which results, altogether too many variables are involved to make this worthwhile. Such variable include salinity and resistivity of the ground or water over which the waves propagates and the water vapour content of the air. The normal procedure is to estimate signal strength with the aid of the tables and graphs available.

**VLF Propagation**

When propagation is over a good conductor like sea water, particularly at frequencies below about 100kHz, surface absorption is small, and so is attenuation due to the atmosphere. Thus the angle of tilt is the main determining factor in the long-distance propagation of such waves. The degree of tilt depends on the distance form the antenna in wavelengths, and hence the early disappearance of the surface wave in HF propagation. Conversely, because of the large wavelengths of VLF signals, waves in this range are able to travel long distances before disappearing (right around the globe if sufficient power is transmitted).

At distances up to 1000 km, the ground wave is remarkably steady, showing little diurnal, seasonal or annual variation. Farther out, the effect of the E layer's contribution to propagation are felt. Also, both short-and long-term signal strength variations take place, the latter including the 11-year solar cycle . The strength of low-frequency signals changes only very gradually, so that rapid fading does not occur. All in all, transmission at these wavelength proves a very reliable means of communication over long distances.

The most frequent users of long-distance VLF transmissions are ship communications, and time and frequency transmissions. Ships use the frequencies allocated to them, from 10 to 110 kHz., for radio navigation and maritime mobile communications. The time and frequency transmissions operate at frequencies as low as 16 kHz (GBR, Rugby, United Kingdom) and 17.8 kHz (NAA, Cutler, Maine). They provide a worldwide continuous hourly transmission of stable radio frequencies, standard time intervals, time announcements, standard musical pitch, standard audio frequencies and radio propagation notices. Such services are also duplicated at HF, incidentally, by stations such as WWV (Ft. Collins, Colorado) and WWVH (Hawaii) operating at 2.5 MHz and the first five harmonics of 5 MHz.

Since VLF antennas are certain to be inefficient, high powers and the tallest possible masts are used. Thus we find powers in excess of 1 MW transmitted as a rule, rather than an exception. For example, the U.S. Naval Communications Station at North-West Cape (Western Australia) has an antenna farm consisting of 13 very tall masts, the tallest 387m height the lowest transmitting frequency is 15 kHz.

**Sky-Wave Propagation-The Ionosphere**

Even before Sir Edward Appleton's pioneering work in 1925, it had been suspected that ionization of the upper parts of the earth's atmosphere played a part in the propagation of radio waves, particularly at high frequencies. Experimental work by Appleton showed that the atmosphere receives sufficient energy from the sun for its molecules to split into positive and negative ions. They remain thus ionized for long periods of time. He also showed that there were several layers of ionization at differing heights, which (under certain conditions) reflected back to earth the high-frequency waves that would otherwise have escaped into space. The various layer, or strata, of the ionosphere have specific effects on the propagation of radio waves, and must now be studied in detail.

***The ionosphere and its effects***

The ionosphere is the upper portion of the atmosphere, which absorbs large quantities of radiant energy from the sun, thus becoming heated and ionized. There are variations in the physical properties of the atmosphere, such as temperature, density and composition. Because of this and the different types of radiation received, the ionosphere tends to be stratified, rather than regular, in its distribution. The most important ionizing agents are ultraviolet and $\alpha$, $\beta$ and $\gamma$ radiation from the sun, as well as cosmic rays and meteors. The overall result, as shown in Fig. 31.8, is a range of four main layers, D, E, $F_1$ and $F_2$, in ascending order. The last two combine at night to form one single layer.

The D layer is the lowest, existing at an average height of 70 km, with an average thickness of 10 km. The degree of its ionization depends on the altitude of the sun above the horizon, and thus it disappears at night. It is the least important layer from the point of view of HF propagation. It reflects some VLF and LF waves and absorbs MF and HF waves to a certain extent.

The E layer is next in height, existing at about 100 km, with a thickness of perhaps 25 km. Like the D layer, it all but disappears at night; the reason for these disappearances is the recombination of the ions into molecules. This is due to the absence of the sun (at night), when radiation is consequently no longer received. The main effects of the E layer are to aid MF surface-wave propagation a little and to reflect some HF waves in daytime.

The $E_s$ layer is a thin layer of very high ionization density, sometimes making an appearance with the E layer. It is also called the sporadic E layer; when it does occur, it often persists during the night also. On the whole, it does not have an important part in long-distance propagation, but it sometimes permits unexpectedly good reception. Its causes are not well understood.

The $F_1$ layer, as shown in Fig. 31.8, exists at a height of 180 km in daytime and combines with the $F_2$ layer at night; its daytime thickness is about 20 km. Although some HF waves are reflected from it, most pass through to be reflected from the $F_2$ layer. Thus the main effect of the $F_1$ layer is to provide more absorption for Hf waves. Note that the absorption effect of this and any other layer is doubled, because HF waves are absorbed on the way up and also on the way down.

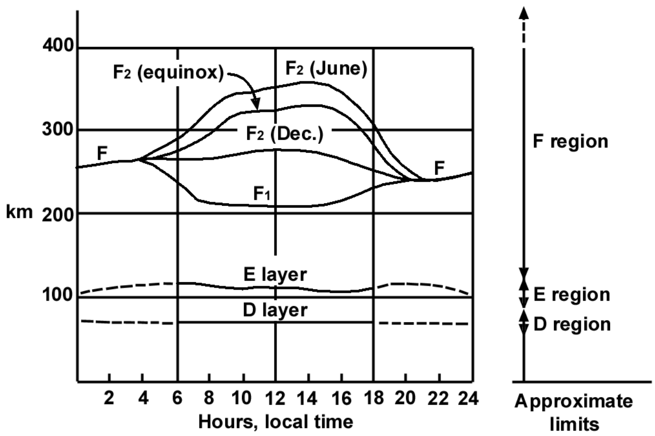



*Fig. 31.8, Ionospheric layers and their regular variations. (F.R. East, "The Properties of the Ionosphere Which Affect HF Transmission")*

The $F_2$ layer is by far the most important reflecting medium for high-frequency radio waves. Its approximate thickness can be up to 200 km, and its height ranges from 250 to 400 km in daytime. At night it falls to a height of about 300 km, where it 250 to 400 km in daytime. At night it falls to a height of about 300 km, where it combines with the $F_1$ layer. Its height and ionization density vary tremendously, as Fig. 31.8, shows. They depend on the time of day, the average ambient temperature and the sunspot cycle (see also the following sections dealing with the normal and abnormal ionospheric variations). It is most noticeable that the F layer persists at night unlike the others. This arises from a combination of reasons; the first is that since this is the topmost layer, it is also the most highly ionized, and hence there is some chance for the ionization to remain at night, to some extent at least. The other main reason is that although ionization density is high in this layer, the actual air density is not, thus most of the molecules in it are ionized. Furthermore, this low actual density gives the molecules a large mean free path ( the statistical average distance a molecule travels before colliding with another molecule). This low molecular collision rate in turn means that, in this layer, ionization does not disappear as soon as the sun sets . Finally, it must be mentioned that the reasons for better HF reception at night are the combination of the $F_1$ and $F_2$ layers into one F layer, and the virtual disappearance of the other two layers, which were causing noticeable during the day.

***Reflection mechanism***

Electromagnetic waves returned to earth by one of the layers of the ionosphere appear to have been reflected. In actual fact the mechanism involved is refraction. As the ionization density increases for a wave approaching the given layer at an angle, so the refractive index of the layer is reduced. Hence the incident wave is gradually bent farther and farther away from the normal.

If the rate of change of refractive index per unit height (measured in wavelength) is sufficient, the refracted ray will eventually becomes parallel to the layer. It will then be bent downward, finally emerging from the ionized layer at an angle equal to the angle of incidence. Some absorption has taken place, but the wave has been returned by the ionosphere (well over the horizon if an appropriate angle of incidence was used).

***Terms and definitions***

The terminology that has grown up around the ionosphere and sky-wave propagation includes several names and expressions whose meanings are not obvious. The most important of these terms will now be explained.

*The virtual height* of an ionospheric layer is best understood with the aid of Fig. 31.9. This figure shows that as the wave is refracted, it is bent down gradually rather than sharply. However, below the ionized layer, the incident and refracted rays follow paths that are exactly the same as they would have been if reflection had taken place from a surface located at greater height, called the virtual height of this layer. If the virtual height of a layer is known, it is then quite simple to calculate the angle of incidence required for the wave to return to ground at a selected spot.

*The critical frequency ($f_c$)* for a given layer is the highest frequency that will be returned down to earth by that layer after having been beamed straight up at it. It is important to realize that there is such a maximum, and it is also necessary to known its value under a given set of conditions since this value changes with these conditions. It was mentioned earlier that a wave will be bent downward provided that the rate of change of ionization density is sufficient, and that this rate of ionization is measured per unit wavelength. It also follows that the closer to being vertical the incident ray, the more it must be bent to be returned to earth by a layer. The result of these two effects is twofold. First, the higher the frequency, the shorter the wavelength, and the less likely it is that the change in ionization density will be sufficient for refraction. Second, the closer to vertical a given incident ray, the less likely it is to be returned to ground . Either way, this means that a maximum frequency must exist, above which rays go through the ionosphere. When the angle of incidence is normal, the name given to this maximum frequency is critical frequency; its value in practice ranges from 5 to 12 MHz for the $F_2$ layer.

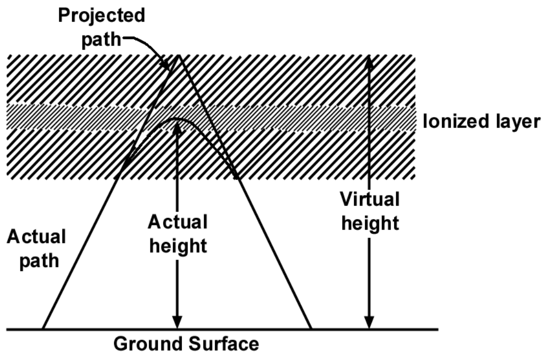



*Fig. 31.9, Actual and virtual heights of an ionized layer.*

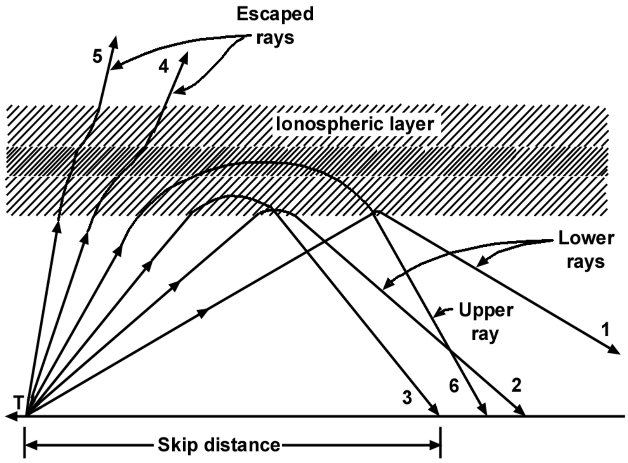



*Fig. 31.10, Effects of ionosphere on rays of varying incidence.*

The maximum usable frequency, or MUF, is also a limiting frequency, but this time for some specific angle of incidence other than the normal. In fact, if the angle of incidence (between the incident ray and the normal) is $\theta$, it follows that

$$MUF = \frac{\text{Critical frequency}}{\cos\theta}$$

$$= f_c \sec\theta \qquad (3)$$

This is the so-called secant law, and it is very useful in making preliminary calculations for a specific MUF. Strictly speaking, it applies only to a flat earth and a flat reflecting layer. However, the angle of incidence is not of prime importance, since it is determined by the distance between the points that are to be joined by a sky-wave link. Thus MUF is defined in terms of two such points, rather than in terms of the angle of incidence at the ionosphere, it is defined as the highest frequency that can be used for sky-wave communication between two given points on earth. It follows that there is a different value of MUF for each pair of points on the globe. Normal values of MUF may range from 8 to 35 MHz, but after unusual solar activity they may rise to as high as 50 MHz. The highest working frequency between a given pair of points is naturally made less than the MUF, but it is not very much less for reasons that will be seen.

The skip distance is the shortest distance from a transmitter, measured along the surface of the earth, at which a sky wave of fixed frequency (more than $f_c$) will be returned to earth. That there should be a minimum distance may come as a shock. One expects there to be a maximum distance, as limited by the curvature of the earth, but nevertheless a definite minimum also exists for any fixed transmitting frequency. The reason for this becomes apparent if the behaviour of a sky wave is considered with the aid of a sketch., such as Fig. 31.10,

When the angle of incidence is made quite large, as for ray 1 of Fig.10, the sky wave returns to ground at a long distance from the transmitter. As this angle is slowly reduced naturally the wave returns closer and closer to the transmitter, as shown by rays 2 and 3. If the angle of incidence is now made significantly less than that of ray 3, the ray will be too close to the normal to be returned to earth. It may be bent noticeably, as for ray 4, or only slightly, as for ray 5. In either case the bending will be insufficient to return the wave, unless the frequency being used for communication is less than the critical frequency (which is most unlikely); in that case everything is returned to earth. Finally, if the angle of incidence is only just smaller than that of ray 3, the wave may be returned, but at a distance farther than the return point of ray 3; a ray such as this is ray 6 of Fig. 31.10, This upper ray is bent back very gradually, because ion density is changing very slowly at this angle. It thus returns to earth at a considerable distance from the transmitter and is weakened by its passage.

Ray 3 is incident at an angle which results in its being returned as closer to the transmitter as a wave of this frequency can be. Accordingly, the distance is the skip distance. It thus follows that any higher frequency beamed up at the angle of ray 3 will not be returned to ground. It is seen that the frequency which makes a given distance correspond to the skip distance is the MUF for that pair of points.

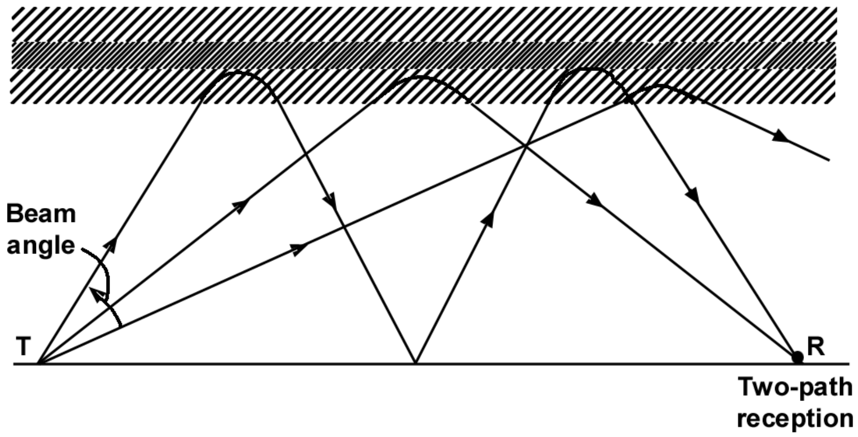



*Fig. 31.11, Multipath sky-wave propagation.*

At the skip distance, only the normal, or lower, ray can reach the destination, whereas at greater distances the upper ray can be received as well, causing interference. This is a reason why frequencies not much below the MUF are used for transmission. Another reason is the lack of directionally of high-frequency antennas. If the frequency used is low enough, it is possible to receive lower rays by two different paths after either one or two hops, as shown in Fig. 31.11, the result of this is interference once again.

The transmission path is limited by the skip distance at one end and the curvature of the earth at the other. The longest single-hop distance is obtained when the ray is transmitted tangentially to the surface of the earth, as shown in Fig. 31.11. For the $F_2$ layer, this corresponds to a maximum practical distance of about 4000 km. Since the semi circumference of the earth is just over 20,000 km, multiple-hop paths are often required, and Fig. 31.12, shows such a situation. No unusual problems arise with multihop north-south paths. However, care must be taken when planning long east-west paths to realize that although it is day "here", it is night "there", if "there" happens to be on the other side of the terminator. The result of not taking this into account is shown in Fig. 31.12 (b). A path calculated on the basis of a constant height of the $F_2$ layer will, if it crosses the terminator, undershoot and miss the receiving area as shown- the F layer over the target is lower than the $F_2$ layer over the transmitter.

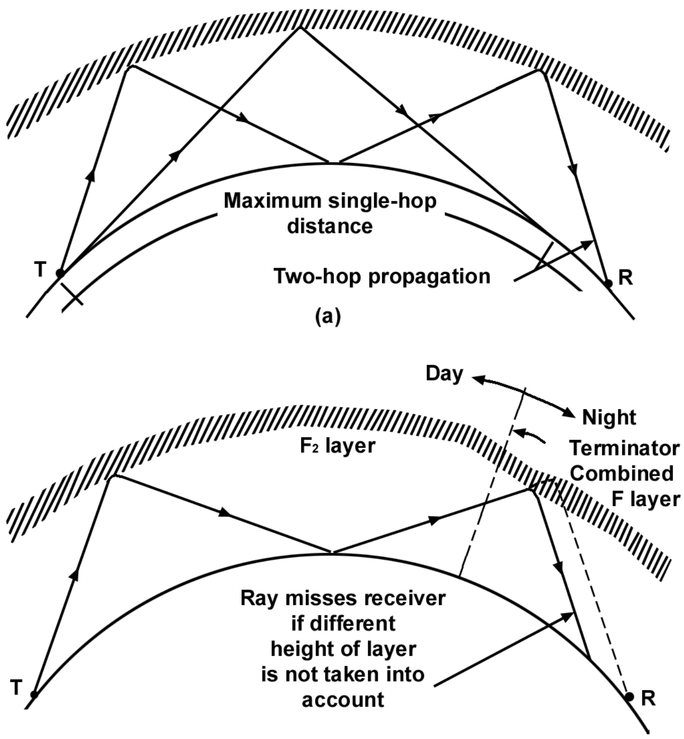



*Fig. 31.12, Long-distance sky-wave transmission paths. (a) North-south; (b) east-west.*

Fading is the fluctuation in signal strength at a receiver and may be rapid or slow, general or frequency-selective. In each case it is due to interference between two waves which left the same source but arrived at the destination by different paths. Because the signal received at any instant is the vector sum of all the waves received, alternate cancellation and reinforcement will result if there is a length variation as large as a half-wavelength between any two paths. It follows that such fluctuation is more as a half-wavelength between any two paths. It follows that such fluctuation is more likely with smaller wavelengths, i.e., at higher frequencies.

Fading can occur because of interference between the lower and the upper rays of a sky wave; between sky waves arriving by a different number of hops or different paths; or even between a ground wave and a sky wave especially at the lower end of the HF band. It may also occur if a single sky wave is being received, because of fluctuations of height or density in the layer reflecting the wave. One of the more successful means of combating fading is to use space or frequency diversity .

Because fading is frequency-selective, it is quite possible for adjacent portions of a signal to fade independently, although their frequency separation is only a few dozen hertz. This is most likely to occur at the highest frequencies for which sky waves are used. It can play havoc with the reception of AM signals, which are seriously distorted by such frequency-selective fading. On the other hand, SSB signals suffer less from this fading and may remain quite intelligible under these conditions. This is because the relative amplitude of only a portion of the received signal is changing constantly. The effect of fading on radiotelegraphy is to introduce errors, and thus diversity is used here wherever possible.

***Ionospheric variations:*** The ionosphere is highly dependent upon the sun, and hence its conditions vary continuously. There are two kinds of variations. The normal ones have already been described as diurnal and seasonal height and thickness changes. Abnormal variations are due mainly to the fact that our sun is a variable star.

The sun has an 11-year cycle over which its output varies tremendously. Most people are unaware of this, because light variations are slight. However, the solar output of ultraviolet rays, coronae, flares, particle radiation and sunspots may vary as much as fiftyfold over that period. The extent of solar disturbance is measured by a method of sunspot counting developed by Wolf in the eighteenth century. On this basis, a pronounced 11-year (± 1 year) cycle emerges, and perhaps also a 90-year supercycle. The highest activities so far recorded were in 1778, 1871 and 1957, which was the highest ever.

The main sun-caused disturbances in the ionosphere are SIDs (sudden ionospheric disturbances, formerly known as Dellinger dropouts) and ionospheric storms. Sudden ionospheric disturbances are caused by solar flares, which are gigantic emissions of hydrogen from the sun. Such flares are sudden and unpredictable, but more likely during peak solar activity than when the sun is "quite". The x-ray radiation accompanying solar flares tremendously increases ionization density, right down to the D layer. This layer now absorbs signals that would normally go through it and be reflected from the F layer. Consequently, long-distance communications disappear completely, for periods up to 1 hour at a time. Studies with earth-based radio heliographs and from satellites in orbit have provided a large amount of data on solar flares, so that some short-term predictions are becoming possible. Two other points should be noted in connection with SIDs. First, only the sunlight side of the earth is affected, and second, VLF propagation is actually improved.

Ionospheric storms are caused by particle emissions from the sun, generally $\alpha$ and $\beta$ rays. Since these take about 36 hours to reach the earth, some warning is possible after large sunspots or solar flares are noticed. The ionosphere behaves erratically during a storm, right around the globe this time, but more so in high latitudes because of the earth's magnetic field. Signal strengths drop and fluctuate quite rapidly. However, using lower frequencies often helps, since the highest ones are the most affected.

Finally, the sporadic E layer previously mentioned is also often included as an abnormal ionospheric disturbance. When present, it has twin effects of preventing long-distance HF communications and permitting over-the-horizon VHF communications. The actual and virtual heights of this layer appear to be the same. This confirms the belief that the layer is thin and dense, so that actual reflection takes place.

Various national scientific bodies have ionospheric prediction programs which issue propagation notices of great value. Among them are the notices of the Central Radio Propagation Laboratory of the United States and the Ionospheric prediction Service of the Australian Department of Science and Technology.

**Space Waves**

Space waves generally behave with merciful simplicity; they travel in (more or less) straight lines ! However, since they depend on line-of-sight conditions, space waves are limited in their propagation by the curvature of the earth, except in very unusual circumstances. Thus they propagate very much like electromagnetic waves in free space. Such a mode of behaviour is forced on them because the ground wave disappears very close to the transmitter, owing to tilt.

***Radio horizon*** **:** The radio horizon for space wave is about four-thirds as far as the optical horizon. This beneficial effect is caused by the varying density of the atmosphere, and because of diffraction around the curvature of the earth. The radio horizon of an antenna is given, with good approximation, by the empirical formula

$$d_t = 4\sqrt{h_t} \tag{4}$$

where $d_t$ = distance from transmitting antenna, km

$h_t$ = height of transmitting antenna above ground , m

The same formula naturally applies to the receiving antenna. Thus the total distance will be given by addition, as shown in Fig.13, and by the empirical formula

$$d = d_t + d_r = 4\sqrt{h_t} + 4\sqrt{h_r} \tag{5}$$

A simple calculation shows that for a transmitting antenna height of 225 m above ground level, the radio horizon is 60 km. If the receiving antenna is 16 m above ground level, the total distance is increased to 76 km. Greater distance between antennas may be obtained by locating them on tops of mountains, but links longer than 100 km are hardly ever used in commercial communications.

***General considerations*** **:** Any tall or massive objects will obstruct space waves, since they travel close to the ground. Consequently, shadow zones and diffraction will result. This is the reason for the need in some areas for antennas higher than would be indicated by Eq.(5). On the other hand, some areas receive such signals by reflection any object large enough to cast a radio shadow will, if it is a good conductor, cause back reflections also. Thus, in areas in front of it a form of interference known as "ghosting" may be observed on the screen of a television receiver. It is caused by the difference in path length (and therefore in phase) between the direct and the reflected rays. This situation is worse near a transmitter than that at a distance, because reflected rays are stronger nearby. Finally, particularly severe interference exists at a distance far enough from the transmitter for the direct and the ground-reflected rays to be received simultaneously.

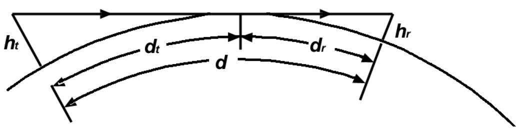



*Fig. 31.13, Radio horizon for space waves.*

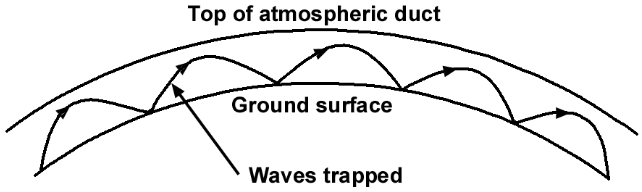



*Fig. 31.14, Super refraction in atmospheric duct.*

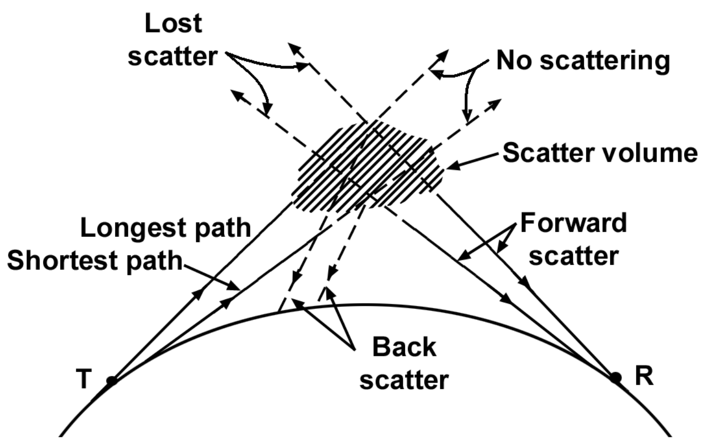



*Fig. 31.15, Tropospheric scatter propagation.*

***Microwave space-wave propagation :*** All the effects so far described hold true for microwave frequencies, but some are increased, and new ones are added. Atmospheric absorption and the effects of precipitation, as already discussed, must be taken into account. So must the fact that at such short wavelengths everything tends to happen very rapidly. Refraction, interference and absorption tend to be accentuated. One new phenomenon which occurs is super refraction, also known as ducting.

As previously discussed, air density decreases and refractive index increases with increasing height above ground. The change in refractive index is normally linear and gradual, but under certain atmospheric conditions a layer of warm air may be trapped above cooler air, often over the surface of water. The result is that the refractive index will decrease far more rapidly with height than is usual. This happens near the ground, often within 30 m of it. The rapid reduction in refractive index (and therefore dielectric constant) will do to microwaves what the slower reduction of these as quantities, in an ionized layer, does to HF waves; complete bending down takes place, as illustrated in Fig. 31.14, Microwaves are thus continuously refracted in the duct and reflected by the ground, so that they are propagated around the curvature of the earth for distances which sometimes exceed 1000 km. The main requirement for the formation of atmospheric ducts is the so-called temperature inversion. This is an increase of air temperature with height, instead of the usual decrease in temperature of 6.5$^o$ C/km in the "standard atmosphere". Super refraction is, on the whole, more likely in subtropical than in temperate zones

**Tropospheric Scatter Propagation**

Also known as troposcatter, or forward scatter propagation, tropospheric scatter propagation is a means of beyond-the-horizon propagation for UHF signals. It uses certain properties of the troposphere, the nearest portion of the atmosphere (within about 15 km of the ground).

***Properties :*** As shown in Fig. 31.15, two directional antennas are pointed so that their beams intersect midway between them, above the horizon. If one of these is a UHF transmitting antenna, and the other a UHF receiving one, sufficient radio energy will be directed toward the receiving antenna to make this a useful communication system. The reasons for the scattering are not fully understood, but there are two theories. One suggests reflections from "blobs" in the atmosphere, akin to the scattering of a searchlight beam by dust particles, and the other postulated reflection from atmospheric layers. Either way, this is a permanent state of affairs, not a sporadic phenomenon. The best frequencies, which are also the most often used, are centered on 900, 2000 and 5000 MHz. However, even here the actual proportion of forward scatter to signals incident on the scatter volume is very tiny between 60 and 90 dB, or one-millionth to one-billionth of the incident power; high transmitting powers are obviously needed.

***Practical considerations* :** Although forward scatter is subject to fading, with little signal scattered forward, it nevertheless forms a very reliable method of over-the-horizon communication. It is not affected by the abnormal phenomena that afflict HF sky-wave propagation. Accordingly, this method of propagation is often used to provide long-distance telephone and other communications links, as an alternative to microwave links or coaxial cable over rough or inaccessible terrain. Path links are typically 300 to 500 km long.

Tropospheric scatter propagation is subject to two forms of fading. The first is fast, occurring several times per minute at its worst, with maximum signal strength variations in excess of 20 dB. It is often called Rayleigh fading and is caused by multi path propagation. As fig. 31.15, shows, scattering is from a volume, not a point, so that several paths for propagation exist within the scatter volume. The second from of fading is very much slower and is caused by variations in atmospheric conditions along the path.

It has been found in practice that the best results are obtained from troposcatter propagation if antennas are elevated and then directed down toward the horizon. Also, because of the fading problems, diversity systems are invariably employed, with space diversity more common than frequency diversity. Quadruple diversity systems are generally employed, with two antennas at either end of the link (all used for transmission and reception) separated by distances somewhat in excess of 30 wavelengths .

**Extraterrestrial Communications**

The most recent, and by far the fastest-growing, field of communications involves the use of various satellite relays, of which the first was launched in 1957, 12 years after the practicability and orbital positioning of stationary satellites were first described and calculated. The field may be subdivided into three parts, each with somewhat differing requirements. First, there is communication with, and tracking of, fast moving satellites in close orbits, typically 145 km in radius. Then there is communication via the geostationary satellites. Such satellites are placed in equatorial orbits at a height of approximately 36,000 km. This height gives a satellites the same angular velocity as the earth, as a result of which it appears to be stationary over a fixed spot on the equator.

***Transionospheric space-wave propagation :*** The ionosphere not only permits long-distance HF propagation; it also affects the propagation of waves through it. Fig. 31.16 showed that those waves which were not reflected by the ionosphere did, nevertheless, suffer varying degrees of bending away from their original paths. This suggests immediately that frequencies used here will need to be well above normal critical frequencies to minimize their refraction. If this were not done, serious tracking errors and communication difficulties would result because of the bending of radio waves. Since refraction becomes insignificant at frequencies above about 100 MHz, while atmospheric absorption is negligible up to about 14 GHz, these two consideration dictate the frequency range used in practice.

A problem encountered in transionospheric propagation is the Faraday effect. This causes the polarization of the radio wave to rotate as it passes through the ionosphere and is a complex process involving the presence of ionized particles and the earth's magnetic field. As the ion density is variable, so is the Faraday effect upon any particular transmission, and so it is not practicable to calculate its extent and make appropriate allowances. If nothing at all is done, severe trouble will result. For instance, a horizontally polarized antenna will receive virtually no energy from a wave which was horizontally polarized when it left the antenna but has become vertically polarized through rotation in the ionosphere.

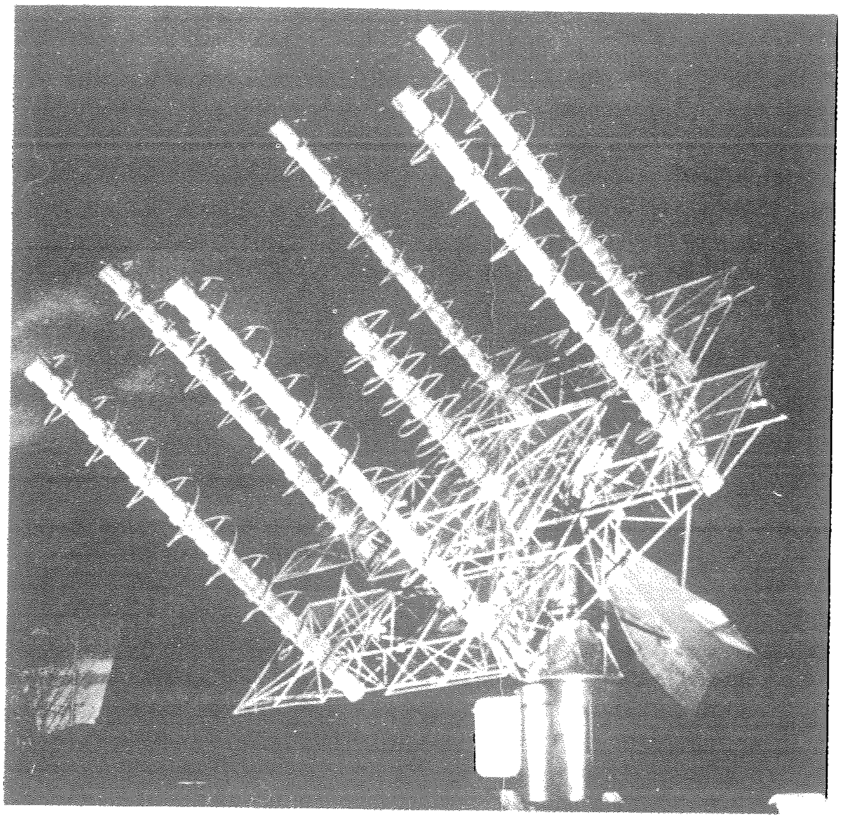


*Fig. 31.16, Circularly polarized antenna array for tracking satellites in low orbits. (Courtesy of Rohe and Schwartz, Munich.)*

Fortunately, several solutions of the problem are possible. One is to use an antenna with circular polarization as shown in Fig. 31.16, and described in detail whose polarization is divided evenly between the vertical and horizontal components. Such an antenna is employed for transmission and reception at the ground station. Thus transmissions to and from the satellite may be received satisfactorily, no matter how they been rotated in the ionosphere. It should be added that such rotation may well exceed 360° at some of the lower frequencies used. Another cure consists in using frequencies above about 1 GHz, at which Faraday rotation is negligible. The present trend for space communications is to frequencies between 1 and 14 GHz partly for the above reason.

***Satellite and probe tracking :*** The requirements for tracking and communicating with satellites in close orbits involve the use of fast-rotating, circularly polarized antennas (such as the array shown in Fig. 31.16), together with fairly low-noise receiver and medium-power transmitters. These are more or less conventional in design and operate in the allocated frequency band from 138 to 144 Mhz. Most of the communication is radiotelemetry since the function of the majority of these satellites is the gathering of scientific and other data. Because these satellites are in low orbits and therefore circumnavigate the earth in about 90 minutes, not only must the ground antennas be capable of fast rotation but also there must be strings of them around the globe because satellites disappear over the horizon in mere minutes.

Tracking interplanetary probes, such as the pioneer and Voyager probes, is a problem of an entirely different order, especially since the distances involved may be several orders of magnitude greater. When it is considered that Voyager 1 and 2 communicated with the earth from Saturn on a power of 30 W from about 1.5 billion km away it can be seen immediately that the first requirements here are for huge directional antenna and extremely low noise receivers as a matter of interest the signal power received by the antennas of NASA's Deep Space Network was only of the order of $10^{-17}$ W. These requirements are identical to those of radio telescopes, which in fact are sometimes used for tracking deep-space probes.

The antennas used, being parabolic reflectors or horns, with diameters often in excess of 60 m. Since the relative motion of a probe is very slow, only the rotation of the earth need be taken into account. The antennas thus have equatorial mountings with motor drives to keep them pointed at the same spot in the sky as the earth rotates. The National Aeronautics and Space Administration uses several such antennas around the world, including one at Goldstone, California, and another near Canberra, Australia.

**SPECTRUM OF TV CHANNEL**

After these preliminaries, we may now turn to the question of transmitting the video signals required for the proper reception of television, noting that the bandwidth occupied by such signals is at least 4 MHz. Bearing in mind filter characteristics, a transmitted bandwidth of 9 MHz would be the minimum required if video transmissions used A3E (which, for that precise reason, they do not). The use of some form of SSB is clearly indicated here to ensure spectrum conservation. So as to simplify video demodulation in the receiver, the carrier is, in practice, sent undiminished. Because the phase response of filters, near the edges of the flat passband, would have a harmful effect on the received video signals in a TV receiver, a portion of the unwanted (lower) sideband must also be transmitted. The result is vestigial sideband transmission, or C3F, as shown in Fig. 31.17. Please note that the frequencies shown there, like the ones used in text, refer strictly only to the NTSC TV system in use in the United States, Canada and Japan. The principles are the same, but the frequencies are somewhat different in the PAL TV system as used in Europe, Australia and elsewhere, and are again different in the French SECAM system.

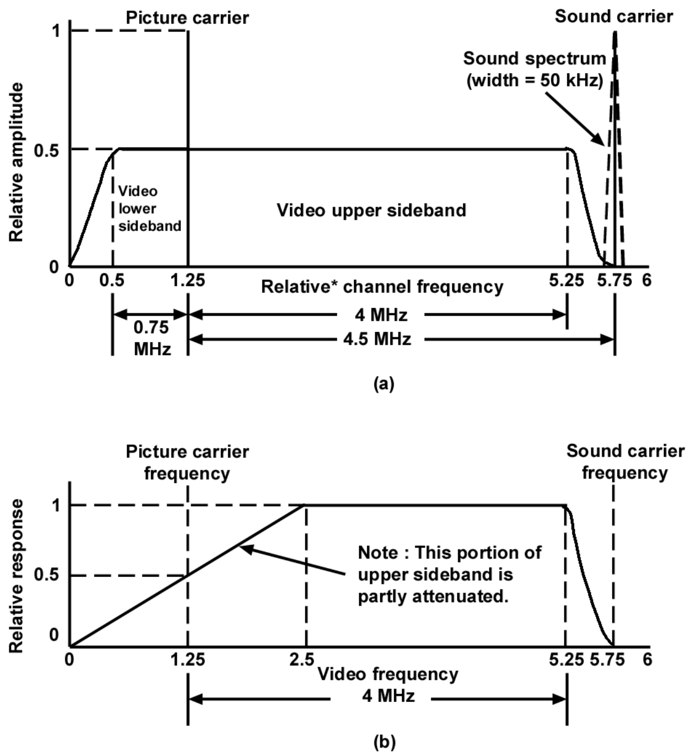



*Fig. 31.17, Vestigial sideband as used for TV video transmission. (a) Spectrum of transmitted signals (NTSC); (b) corresponding receiver video amplifier frequency response.*

By sending the first 1.25 MHz of the lower sideband (the first 0.75 MHz of it undiminished), it is possible to make sure that the lowest frequencies in the wanted upper sideband are not distorted in phase by the vestigial - sideband filter. Because only the first 1.25 MHz of the lower sideband is transmitted, 3 MHz of spectrum is saved for every TV channel. Since the total bandwidth requirement of a television channel is now 6 MHz instead of 9 MHz, clearly a great saving has been made, and more channels consequently can be accommodated.

For completeness, Fig. 31.17 a shows also the location, in frequency, of the frequency-modulated sound transmissions that accompany the video. It should be noted that these transmissions have nothing to do with the fact that the modulation system for video is C3F, and would have been there regardless of the video modulation system. All these signals occupy frequencies near the video transmissions simply because sound is required with the pictures, and it would not be very practical to have a completely separate receiver for the sound, operating at some frequency remote from the video transmitted frequencies.

Fig. 31.17 b shows the video frequency response of the television receiver Attenuation is, as can be seen, purposely provided for the video frequencies from 0 to 1.25 MHz. The reason is quite simple: Extra power is transmitted at these frequencies (since they are sent in both sidebands, whereas the remaining video frequencies are not only in the upper sideband). Accordingly, these frequencies would be unduly emphasized in the video output of the receiver if they were not attenuated appropriately.

# CHAPTER : 32
# COMMUNICATIONS SYSTEM

## GENERAL

Communications and aviation are the two major functions of airborne radio. Communication systems primarily involve voice transmission and reception between aircraft to aircraft and ground stations. Radios are used in aircraft as navigational aids in a number of applications. They range from a simple radio direction finder to navigation system which use computers and other advanced electronic techniques to automatically solve the navigational problems for an entire flight. Marker becon receivers, instruments landing systems. (Involving radio signals for glide slope and direction), distance measuring equipments, radar, area navigation systems, and omnidirectional radio receivers are but a few basic application of air borne radio navigational system available for installation and use in aircraft. Safe Aircraft operation is dependent to a large degree upon the satisfactory performance of the air borne. Communications and navigation system.

## BASIC RADIO PRINCIPLES

The principle of radio communication can be shown by using a simple transformer. As shown in figure 32.1, closing the switch in the primary circuit causes the lamp in the secondary circuit to be illuminated opening the switch extinguishes the light.

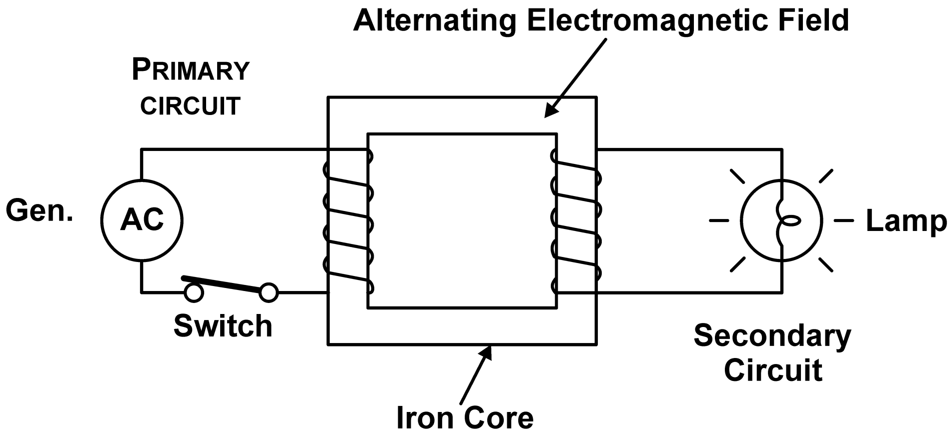



Fig. 32.1, Basic radio principle.

There is no direct connection between the primary and secondary circuit. The energy that illuminates the light transmitted by an aircraft alternating electromagnetic field, in the core of the transformer. This is simple form of wireless control of our circuit by another.

The basic concept of radio communications involves the transmission and reception of electromagnetic (radio) energy wave through space. A.C. passing through a conductor creates electromagnetic fields around the conductor, Energy is alternately stored in these fields and returned to the conductor. As the frequency of current alternation increases, less and less of the energy stored in the field return to the conductor. Instead of returning the energy is radiated in space in the form of electromagnetic waves. A conductor radiating in this manner is called antenna.

For an antenna to radiate efficiently, a transmitter must supply it with an aircraft of the selected frequency. The frequency of the radio wave radiated will be equal to the frequency of the applied current when current flows through a transmitting antenna, radio waves are radiated in all direction in much the same way as wave travels on the surface of a panel into which a rock has been thrown. Radio waves travels at a speed approximately 186000 miles/sec.

If a radiated electromagnetic field passes through a conductor some of the energy in the field will set electrons in motion in the conductor. This electron flow constitutes a current that varies with changes in the electromagnetic field. Thus a variation of the current in a radiating antenna causes a similar varying current in the conductor (receiving antenna) at a distant location. Any intelligence being produced as current in a transmitting antenna will be reproduced as current in a receiving antenna.

## FREQUENCY BANDS

The radio frequency portion of the electromagnetic spectrum extends from approximately 30 kHz to 30,000 MHz. As a matter of convenience, this part of the spectrum is divided into frequency bands. Each band or frequency range produced different effects in transmission. The radio frequency bands proven most useful and presently in use are :-

| Frequency | Range | Bands |
|---|---|---|
| Low frequency | (L/F) | 30 to 300 kHz |
| Medium frequency | (M/F) | 300 to 3000 kHz |
| High frequency | (H/F) | 3000 KHz to 30 MHz |
| Very High frequency | (VHF) | 30 to 300 Mhz |
| Ultra High frequency | (UHF) | 300 to 3000 MHz |
| Super High frequency | (SHF) | 3000 to 30,000 MHz. |

In practice, radio equipments usually covers only a portion of the designed band e.g. civil VHF equipment normally operates at frequencies between 108 MHz and 135.95 MHz.

**BASIC EQUIPMENT COMPONENTS**

The basic components (Fig. 32.2)of a communication system are; microphone, transmitter, transmitting antenna, receiving antenna, receiver and a head set or loud speaker.

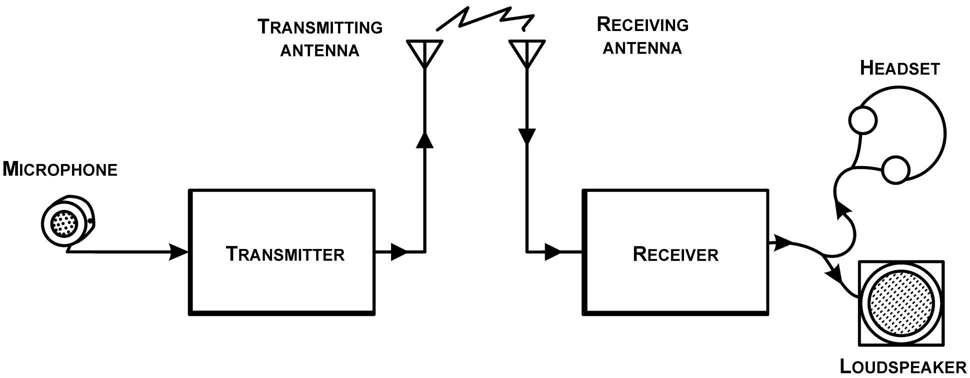



Fig. 32.2, Basic communication system.

**Transmitters**

The transmitter may be considered as a generator which changes electrical power into radio waves. A transmitter must perform these functions (1) generate a RF (radio frequency) signal. (2) amplify the RF signal and (3) provide a means of placing intelligence on the signal.

The transmitter contains an oscillator circuit to generate the RF signal (or a sub harmonic of the transmitter frequency ,if frequency doubles or multipliers are used) and amplifier circuits to increase the output of the oscillator to the power level required for proper operation.

The voice (audio) intelligence is added to the RF signal by a special circuit called the modulator. The modulator uses the audio signals to vary the amplitude or frequency of the RF signal. If the amplitude is varied the process is called amplitude modulation or AM. If the frequency is varied , the process is known as frequency modulation or FM.

Transmitters take many forms, have varying degrees of complexity and develop various levels of power. The amount of power generated by a transmitter affects the strength of the electromagnetic field radiating from the antenna. Thus it focuses that the higher, the power output from a transmitter, the greater the distance its signal may be received.

VHF transmitters used in single engine and light twin engine aircraft vary in power output from 1 watt to 30 watts, depending on the particular model radio. However, radios having 3 to 5 watt rating are most frequently used. Executive and large aircraft are usually equipped with VHF transmitter having a power output of 20 to 30 watts.

Aviation Communication transmitters are crystal controlled in order to meet the frequency to clearance requirements of F.C.C. Most transmitters are selected for more than one frequency. The frequency of channel selected is determined by a crystal transmitters may have from one to 680 channels.

**Receivers**

The communications receiver must select radio frequency signals and convert the intelligence contained on these signals into a usable form; either audible signals for communication and audible or visual signal for aviation.

Radio waves of many frequencies are present in the air. A receiver must be able to select the derived frequency from all those present and amplify the small aircraft signal voltage.

The receiver contains a demodulator circuit to remove the intelligence. If the demodulator circuit is sensitive to amplitude changes, it is used in AM sets and called a detector. A demodulator circuit that is sensitive to frequency changes is used for FM reception and is known as a discriminator.

Amplifying circuits within the receiver increases the audio-signal to a power level which will operate the headset or loud speaker properly.

**Antenna**

An antenna is a special type of electrical circuit designed to radiate and receive electromagnetic energy. As mentioned previously, a transmitting antenna is a conductor which radiates electro magnetic waves when a radio frequency current is passed through it. Antennas vary in shape and design (Fig. 32.3) depending upon the frequency to be transmitted, and special purposes they must serve. In general, communication transmitting stations radiate signals in all directions. However special antennas are designed that radiate only in certain directions or certain beam patterns.

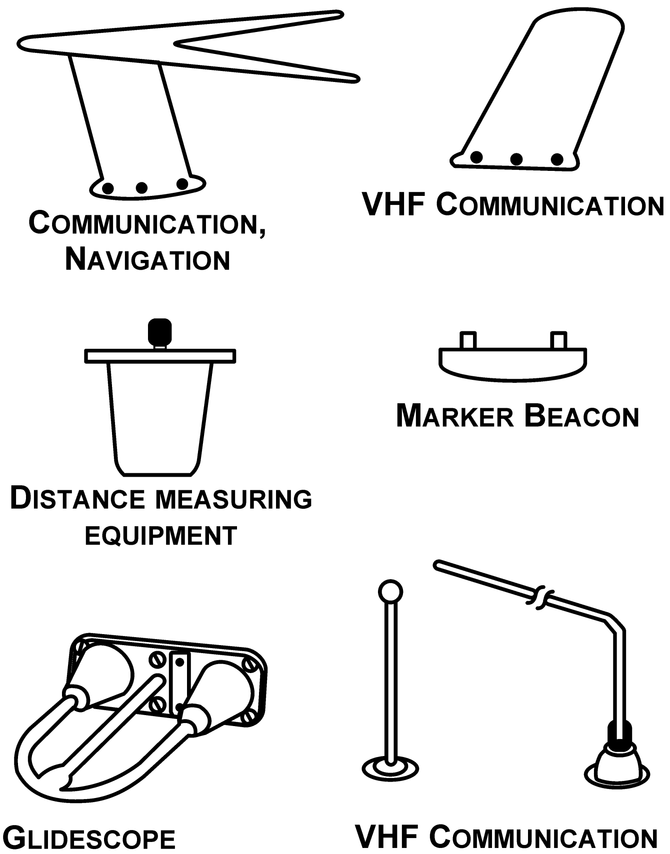



*Fig. 32.3, Various types of antennas.*

The receiving antenna must intercept the electromagnetic waves that are present in the air. The shape and size of the receiving antenna will also vary according to the special purpose for which it is intended. In air borne communications the same antenna is normally used for both transmission and reception of signals.

**Microphones**

A microphone is essentially an energy convertor that changes acoustical (sound) into corresponding electrical energy. When spoken into a microphone, the audio pressure waves generated strike the diaphragm of the microphone causing it to move in and out in accordance with the instantaneous pressure delivered to it. The diaphragm is attached to a device that causes current to flow in proportion to the pressure applied.

For good quality sound the electric waves from a microphone must correspond closely in magnitude and frequency to the sound waves that causes them, so that no new frequencies are introduced.

A desirable characteristic is the ability of the microphone to favour sounds coming from a nearly source over random sounds coming from a relatively greater distance. When taking into this type of microphone, the lips must be held as close as possible to the diaphragm.

Readable radio transmission depends on the following factors :- (i) Voice amplitude (ii) rate of speed (iii) pronunciation and phrasing. When using a microphone, speak loudly, without existing extreme effort . Talk slowly enough so that each word is spoken distinctly. Avoid using unnecessary words.

**Power Supply**

The power supply is a component that furnishes the correct voltage and current needed to operate the communication equipment. The power supply can be a separate component or it may be contained with in the equipment it supplies. Electromechanical devices used as electronic power supplies include dynamotors and inverters.

The dynamotor performs the dual functions of motor and generator, changing the relatively low voltage of the aircraft electrical system into a much higher value. The multi vibrator is another type of voltage supply used to obtain in high a.c. and d.c. voltage from a comparatively low d.c. voltage.

In many aircraft, the primary source of electric power is direct current. An inverter is used to supply the required A.C. common aircraft inverter consist of a d.c. motor driving an a.c. generator. Static, or solid state inverters are replacing the electro mechanical inverters in many applications. Static inverters have no moving parts but use semi conductors devices and circuits that periodically pulse d.c. current through the primary of a transformer to obtain an a.c. output from the secondary.

**COMMUNICATION SYSTEMS**

The most common communication system in use today is the VHF system. In addition to VHF equipment; large aircraft are usually equipped with HF communication system. Air borne communications system vary considerably in size Wt. power requirement, quality of operation and cost depending upon the desired operation.

Many air borne VHF and HF. communication system are transceivers. A transceiver is a self contained transmitter and receiver which share common circuits i.e. power supply, antenna and tuning. The transmitter and receiver both operate on the same frequency, and the microphone button determines when there is an output from the transmitter. In the absence of transmission, the receiver is sensitive to in coming signals. Since weight and space are of great important in aircraft, the transceiver is widely used. Large aircraft may be equipped with transceivers or a communication system that uses separate transmitter and receivers.

**VHF (Very High Frequency) Communications**

VHF air borne communication sets operate in the frequency range from 108 MHZ to 135.95 MHZ. VHF receivers are manufactured that covers only the communications frequencies, or both communications and navigational frequencies. In general the VHF radio waves follow approx. straight lines. Theoretically, the range of contact is the distance to the horizon and this distance is determined by the height of the transmitting and receiving antennas. However, communications is sometimes possible many hundreds of miles beyond the assumed horizon range.

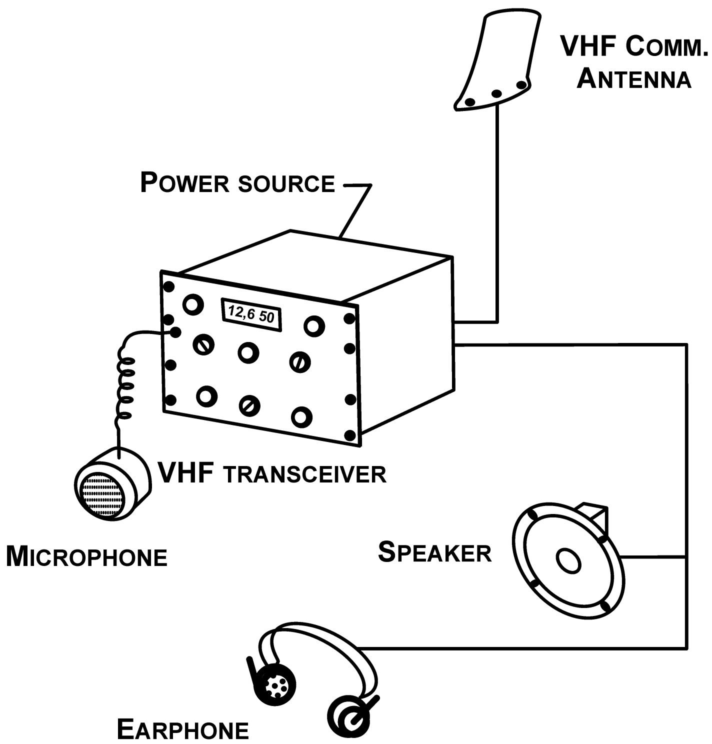



*Fig. 32.4, VHF system diagram.*

Many VHF radio have the transmitter receiver, power supply and operating controls units into a single unit. This unit is frequently installed in a cutout in the instrument panel. A system diagram of a typical panel mounted VHF transceiver is shown in fig. 32.4 other have certain portions of the communication system mounted on the instrument panel and remainder remotely installed in a radio or luggage compartment.

To perform an operational check of a VHF communication system, a source of electric power must be available. After turning the radio control switch "on" allow sufficient time for the equipment to warm up before beginning the operational checks. Using the frequency selector, select the frequency of the ground station to be contacted. Adjust the volume control to the desired level.

With the microphone held close to the mouth, press the microphone button and speak directly into the microphone to transmit, when through talking release the button. This action will return the communication receiver to operation when the ground station acknowledge the initial transmission request that operation check be made on all frequencies or channels. Prior to transmitting make certain that a station licence is displayed in the aircraft.

**HF (High Frequency) Communications**

A high frequency communication system is used for long range communication. HF system operate essentially in the same as VHF system, but operate in the frequency range from 3 MHz to 30 MHZ. Communications over long distances are available with HF radio because of longer transmission range HF transmitters have higher power outputs than VHF transmitters.

The design of antennas used with HF communication systems vary with the size and shape of the aircraft. Aircraft which fly (Guide) below. 300 m.p.h. generally use a long wire antenna. Higher speed aircraft have specially designed antenna probes installed in the vertical stabilizer. Regardless of the type antenna, a tuner is used to match the impedance of the transceiver to the antenna.

An operational check of an HF radio consists of turning the control switch to "on" adjusting the RF gain and volume controls, selecting the derived channel and transmitting the appropriate message to the called station. Best adjustment of the gain control can be obtained with the volume control set at half range. The gain control is used to provide the strongest signal with the least amount of noise. The volume control is used to set sound level and affects only the loudness of the signals.

■■■

# INDEX


## A

A simple model of a computer. 288
A.C. measuring instruments 138
A.C. Synchronous Systems 188
A.C. motors 84
Addition/Subtraction in 2's Complement Representat 285
Advantages 88, 234, 273
Advantages and Disadvantages 239, 271
Advantages of Negative Feedback 274
Advantages of RC Coupling 270
Advantages of Transformer Coupling 272
Air Filter 169
Aircraft Batteries 101
Aircraft Electrical Test Equipments 128
Aircraft Instrument Panels and Range Marking 155
Aircraft radio navigation instruments 208
Aircraft Sighting Points 177
Air-Oil Separator 168
Airspeed Indicators 165
Alphanumeric Displays 225
Alternating current generators 72
Altimeters 163
Ammeter 128
Amplifier Classification Based on Biasing Conditio 267
Amplifier Coupling 269
Antenna 209, 211, 312
Application 241, 242
Applications 87, 88, 89, 231, 239, 243, 273
Armature Reaction in DC Generators 62
Armature Core 53
Armature Windings 54
Assumed Direction of Current 33
Auto pilot 214
Auto Transformers 123
Automatic Direction Finder (ADF) 209
Automatic Direction Finder (ADF) 209
Automatic Direction Finder Equipment 209
Autopilot system 214
Average Values 232, 233, 234

## B

BACK E.M.F. 78
Back Pitch (YB) 56
Bandwidth and required spectra 297
Base 248
Basic autopilot components 215
Basic Desynn Circuit 187
Basic Equipment components 311
Basic Gauge System 191
Basic radio principles 310
Batteries Received Dry and Uncharged 104
Batteries Received with Electrolyte 104
Battery Cleaning 111
Battery records 107, 117
Bearing Indicator 209
Biasing the Diode 239
Binary Addition 283
Binary arithmetic 282
Binary Division 284
Binary Multiplication 284
Binary number system 277
Binary Substraction 284
Binary-to-Decimal Conversion 277
Bipolar Junction Transistor 248
Body-end-dot system 30
Bond Testing 147
Bonding 144
Bonding Carrying the Main Electrical Supply 147
Bonding Connections. 145
Bonding of Aircraft of Metallic and Non-Metallic C 144
Bonding Tester Servicing 148
Branch 5
Brush Shift 66

## C

Cage rotor 84
Capacitance-Type Fuel-Gauge System 188
Capacitor Motors 87
Capacitors In Alternating-Current Circuits 190
Capacitors in Series And Parallel 190
Capacitor-start Motor 87
Capacity Recycling Procedures 113
Capacity Test 113
Capacity tests 105
Capillary Thermometers 199
Capillary Type 197
Carbon Composition 16
Carbon Pile Compression 96
Carbon-pile Voltage Regulator 95
CB Circuit 253
CB Configuration 251
CC Configuration 253
CE Circuit 254
CE Configuration 252
Cell Balancing 114
Cell Removal and Replacement. 114
Cermet (Ceramic Metal) 17
Characteristics 267, 268, 269
Characteristics of A Shunt or Separately Exc 80
Characteristics of a CC Amplifier 266
Characteristics of a CB Amplifier 264
Characteristics of a CE Amplifier 265
Characteristics of A Compound Motor 82
Characteristics of A D.C. Series Motor 81
Characteristics of Compound Dc Generators 70
Characteristics of computers 288
Characteristics of DC Generators 67
Charging Conditions 103
Charging of Batteries 111
Charging of Individual Cells 112
Check on the Direction of Magnetic North 182
Chemical 128
Chemical principle 101
Chemical Principle 109
Circuit ABDA 34
Circuit ADCEA 34
Circuit BCDB 34
Circuit Operation 262, 264, 266, 270, 272
Class-B Amplifier 268
Class-C Amplifier 269
Classification of Amplifiers 262

Classification of Computers 293
Coefficient of Coupling 43
Coefficient of Mutual Inductance (M) 43
Coefficient of Self-induction (L) 42
Coil and Winding Element 54
Coil-span or Coil-pitch (YS) 55
Cold-Junction Temperature Compensation 203
Collector 248
Colour CRT Displays 222
Colour generation 223
Command Elements 216
Common Base (CB) Amplifier 262
Common Base Formulas 259
Common Base Static Characteristics 255
Common Base Test Circuit 255
Common Collector (CC) Amplifier 265
Common Emitter (CE) Amplifier 264
Common Emitter Formulas 260
Common Emitter Static Characteristics 257
Common Emitter Test Circuit 257
Communication Systems 313
Communications System 310
Commutating Poles or Interpoles 66
Commutation 63
Commutator 54
Commutator Pitch (YG) 56
Comparison between Single-phase and Three-phase in 89
Comparison of Cage and Wound Rotors 85
Compass Swinging 185
Compass Calibrator Set 182
Compass Errors and Methods of Compensation 174
Compass Errors and Methods of Compensation 176
Compass Swinging Area 175, 177
Compass Swinging Procedure 175
Compass Swinging Procedures 177
Compensating windings 67
Compensation for Errors 200
Compound motor 80
Compound Wound 61
Computer or Amplifier 217
Conductance and conductivity 15
Conductor 54
Connection to Charging Equipment 112
Constant Frequency Generator 76
Constant Frequency Generator Construction 73
Constant frequency systems 73
Constant-current Characteristic 246
Constant-Current Charging 111
Construction 109
Construction 229, 237, 242
Construction 241
Construction 51, 91, 96
Construction 84
Construction Of Voltage Transformers 121
Continuity Testing 148
Conventional Swinging Procedure 181
Crosstalk Error Compensation 177
Crosstalk Errors 176
Current equation 79
Current Limiting 99
Current Regulator 95
Current Transfer Characteristic 257, 258
Current Transformers 122
Cutoff and Saturation Points 260

## D

D.C. measuring instruments 128
Damping 131
D’Arsonval Meter 128
DC Generator Principle 51
DC motors 78
Decimal Prefixes 275
Decimal to binary conversion and vice -versa 276
Decimal-to-Binary Conversion 278
Definitions 171
De-icing Pressure Gauge 197
Delta/Star* Transformation 6
Deposited Carbon 16
Desynn Torque Pressure Indicating System 194
Determination of Crosstalk Errors 183
Determination of Sign 33
Determining the Value of a Shunt 131
Dials 155
Different Ways of Drawing Schematic Transistor Cir 258
Diode as a Rectifier 231
Diode Parameters 230, 238
Diode Resistance 241
Direct current generator 51
Direct Reading Fuel Quantity Gauge 188
Direct-coupled Two-stage Amplifier 272
Direction Indicators 173
Direction of Induced E.M.F. and Current 39
Direct-Reading Indicators 192
Direct-Reading Magnetic Compasses 174
Disadvantages 272, 273
Distance Measuring Equipment (DME) 210
Division of Current in Parallel Circuits 20
Double-Spiral Bourdon Tube 200
Drains 160
Dynamically Induced E.M.F. 40

## E

E1 and E2 Voltages Check 183
Earth Gyro 171
Earth Terminals 146
Effect of temperature on resistance 15
Effects of Armature Reaction 63
Effects of Current 128
Electric circuits and network theorems 5
Electrical Capacitance 189
Electrical Characteristics of a Triode 245
Electrical Damping 131
Electrical Diagram Symbols 151
Electrical Indication Systems 198
Electrical Leakage Check 113
Electrical Methods 177
Electrical Resistance Thermometers 201
Electrodynamometer Ammeter 139
Electrodynamometer Meter Movement 138
Electrodynamometer Voltmeter 140
Electrolyte Level 112
Electrolyte Level and Adjustments 104, 110
Electromagnetic 128
Electromagnetic induction 38
Electromagnetism 46
Electronic (CRT) Displays 219
Elementary knowledge of computers, its application 288
Emitter 248

Encoding Altimeters 165
Engine Cylinder Head Temperature Indicating System 204
Engine Gauge Unit 195
Engine Instruments 187
Engine Speed Indicators 198
Engine Vibration Indicating Systems 205
Engine-Driven Vacuum Pump 168
Equivalent Circuit 238
Equivalent resistance 24
Explanation 236
Extending the range of an Ammeter 131
Extending the Voltmeter Range 134
External Circuits 204
Extraterrestrial Communications 307

**F**

Factors on Which Capacitance Depends 190
Faraday's Laws of Electromagnetic Induction 39
Feedback Amplifiers 273
Field strength at a distance 300
First Decade (1976-85) 290
First generation of computers 289
Flexible Drive Shafts 198
Flight Instruments : Gyroscopic Systems 167
Flight Instruments : Pitot-Static Systems 158
Float-Type Fuel-Quantity Indicating Systems (Elect 188
For a simplex lap-wound generator 60
For a simplex wave-wound generator 60
Forward Characteristic 229
Free Gyro 171
Frequency bands 310
Frequency meters 142
Frequency Spectrum of the AM Wave 295
Frequency Spectrum of the FM Wave 295
Frequency Wild Generators 75
Frequency wild systems 72
Front Pitch (YF) 56
Fuel Flowmeter Systems 205
Fuel Flowmeters 205
Fuel Quantity Gauges 188
Full-Wave bridge Rectifier 233
Full-wave Rectifier 232
Function of Bonding. 144
Functioning Tests 150

**G**

Gassing 112
Gate Check Valve 168
General 32, 94, 144, 155, 171, 187, 310
General construction 101
General description 109
Generated E.M.F. or E.M.F. Equation of a Generator 60
Generator Construction 72
Ground (Surface) Waves 300
Gyro Horizons 172
Gyroscopes 171
Gyroscopic Inertia 171

**H**

Half-Wave Rectifier 231
HF (High Frequency) Communications 314
High-Voltage Ink Film 16
How to Remember? 6, 7
Hydraulic Pressure Gauge 196

I

Ideal Constant-Current Source 7
Ideal Constant-Voltage Source 7
Impedance coupled Two-stage Amplifier 271
Importance of VCE 261
Important Biasing Rule 249
Inclined-Coil Iron-Vane Meter 140
Index Errors Compensation 177
Index Error 174, 176
Induced E.M.F. 40
Inductances in parallel 45
Inductances in Series 44
Initial Filling 102
Input characteristic 255, 257
Inspection 110
Inspection and testing of Circuits 148
Inspection before Charging 102
Inspection of Wiring System 148
Installation 106, 115, 158, 159
Instantaneous Vertical Speed Indicator 166
Instrument cases 155
Instrument landing system (ILS) 208
Instrument panels 156
Insulation Resistance Test 105, 114
Insulation Resistance Testing 149
Integrated Drive Generators 74
Introduction 101, 109, 143, 276

**K**

Kirchhoff's First Law or Point Law or Current Law 32
Kirchhoff's law 32
Kirchhoff's Laws 32
Kirchhoff's Second Law or Mesh Law or Voltage Law 32

**L**

Lap and Wave Windings 57
Laptop PCs 293
Large Forward Bias 237
Laws of resistance 15
Lead-Acid Battery 101
Leak Testing Pitot-Static Systems 163
Leakage Currents in a Transistor 253
Leakage Test 106
Lenz's Law 40
Limitations 88
Location 174
Logic gates and truth tables 287
Loop 5

**M**

Machmeters 166
Magnet Core Airgap 96
Magnetic Alignment of Detector Unit 182
Magnetic compass 208
Main features of synchronous motor 92
Mainframe Computers 293
Maintenance 101, 110
Maintenance of installed batteries 106, 116

Maintenance of Pitot-static systems 162
Manifold Pressure Gauges 193
Maximum Allowable Air speed Indicators 165
Mechanical Damping 131
Mechanical Indicators 198
Mechanical Methods 177
Megger (Megaohmmeter) 137
Mercury-in-steel thermometer 199
Mesh 5
Metal Film 16
Metal Glaze 17
Meter Sensitivity 131
Methods of Compensation 177
Methods of Improving Commutation 65
Micro-Desynn Circuit 187
Microphones 312
Millivolt Drop test 148
Motor Characteristics 87, 88
Motor Principle 78
Moving Iron-Vane Meter 140
Multimeters 132
Mutual Inductance 43
Mutually-induced e.m.f. 41

**N**

Negative Feedback 274
Nickel Cadmium Battery 109
No. Forward Bias 236
Node 5
Nonlinear resistors 17
Number systems 276

**O**

Observations 296
Occasions for Compass Swinging 185
Ohmmeters 135
Ohm's law 9
One -Cycle Error Compensation 177
One-Cycle Errors 174, 176
One's Complement Representation 280
'Opens' in a parallel circuit 19
'Opens' in a series circuit 18
Operating limitations of rectifiers 119
Operation 241, 242
Operation of the Meter Movement 130
Optical Transfer Of Flux Detector Unit 184
Output Characteristic 256
Output Elements 217
Output or Collector Characteristic 257

**P**

P.N Junction Diode 229
Palmtop PCs 293
Percentage Speed Indicators 199
Permanent-split Capacitor (PSC) Motor 88
Personal Computers (PCs) 293
Phase Reversal in Amplifiers 267
Photoelectric 128
Physiological 128
Piezoelectric 128
PIN Diode 241
Pipelines 159
Pitch of a Winding (Y) 55
Pitot-Static System 160
Plate Characteristic of a Triode 246
Plumb Line Sighting 179
Pole Coils 53
Pole Cores and Pole Shoes 52
Pole-pitch 54
Positive Feedback 274
Power 9
Power conversion equipment 118
Power equations 79
Power Indicators For Turboprop Engines 193
Power Loss, Engine Pressure Ratio and Percentage 195
Power Supply 313
Practical Applications 171
Practical Generator 52
Precession 171
Preparation before Swinging 175
Preparations Prior to Test 149
Pressure Error 159
Pressure Heads 158
Pressure Indicators 192
Pressure Relief Valve 168
Pressure Switches 198
Pressure Transmitter System 197
Primary and Secondary Conductors. 144
Principle 120, 203
Principle of Operation-SH-II 96
Principle of Feedback Amplifiers 273
Principle of Operation 92, 100, 214
Principle of Remote - Reading Compass Systems 176
Principle of the CRT 220
Principles 94
Problem-solving using computers 289
Production of Induced E.M.F. and Current 38
Propagation of waves 299

**R**

Radars 210
Range Markings 155
Rate Gyro 171
Rate Gyros 171
Ratiometer Circuit 202
Ratiometer electrical resistance thermometer 202
Ratiometer pressure Indicators 197
RC-coupled Two-stage Amplifier 270
Receivers 311
Receiving Equipment 212
Re-Charging a Battery in Service 104
Rectifier A.C. Meters 138
Rectifiers 118
Reflection mechanism 302
Regulator Adjustments 96
Rejected Batteries or Cells 115
Relation Between a and b 252
Relation Between Magnetism and Electricity 38
Relations Between Transistor Currents 253
Remote-indicating Instruments 197
Remote-reading compasses 176
Resistance 14
Resistance Commutation 65
Resistance Values 146
Resistor colour code 28
Restrictor Valve 169

Resultant Pitch (YR) 56
Reversal of Direction of Rotation 86, 88
Reverse Characteristic 230
Reverse Current Cut-out Relay 99
Rotary Converting Equipment 124
Rotation Indicators 199
Rument Panels and Range MAR 155–157

**S**

Safety Precautions. 101
Scan conversion 223
Schottky Diode 242
Screen format 223
Sealed Batteries 109
Second Phase (1986-2000) 290
Selector Valve 169
Selenium Rectifiers 119
Self-excited 61
Self-induced e.m.f. 42
Self-inductance 42
Semiconductor Diodes 229
Semi-open Batteries 109
Semi-sealed Batteries 109
Sensing Elements 217
Sensor Units 201
Separately Excited DC Generator 67
Separately-excited 61
Series motor 79
Series Wound 61
Series-type Ohmmeters 136
Shaded-pole Motors 88
Short and open circuits 5, 18
Short' in a series circuit 18
Shorts' in parallel circuits 19
Shunt motor 78
Shunt wound 61
Shunt-Type Ohmmeter 136
Sighting Rods 178
Signed binary numbers 280
Sign-Magnitude Representation 280
Silicon Controlled Rectifier (S.C.R.) 119
Silicon Rectifiers 119
Simple Loop Generator 51
Simplex Lap-winding 59
Simplex Wave Winding 58
Single-layer Winding 56
Single-phase Induction Motor Principle 85
Single-phase Series (Universal) Motor 89
Sky-Wave Propagation-The Ionosphere 301
Slab-Desynn circuit 187
Small Forward Bias 236
Solid-state Voltage Regulators 99
Source Conversion 7
Sources of Pitot And Static Pressures 158
Sources of power for gyro operation 167
Space Waves 305
Spectrum of TV channel 308
Spectrum of waves 295
Speed/armature current characteristic 81
Speed/torque characteristic 82
Speed-armature current characteristics 80
Split-phase Induction Motor 86
Star/Delta Transformation 6
State of Charge 105, 112
Static converting equipment 118
Static Inverters 126
Static Vents 158
Statically Induced E. M. F 41
Step recovery diode 243
Stereophonic FM Multiplex System 298
Storage and transportation 107, 117
Subtraction Using 2's Complement 285
Suction 170
Suction Gauge 169
Suction Relief Valve 168
Summing Up 250
Supercomputers 293
Swinging by Inertial Navigation Systems 185
Swinging Procedure 182
Synchro Torque Pressure Indicating System 194
Synchronizing And Indicating Devices 212
Synchronous Data Transmission Systems 187
Synchronous motor 91
Synchroscopes 199

**T**

Telescopic Target Fixture 178
Temperature coefficient of resistance 16
Temperature Indicators 199
Terms and definitions 302
Test Result. 149
Testing the System 149
The ADF Receiver 209
The Applications of Radar 213
The Bipolar junction transistor 248
The Fifth Generation 291
The Fourth Generation 290
The ionosphere and its effects 301
The Second Generation 289
The Third Generation 290
The Unit Of Resistance 14
Theory 243
Thermal 128
Thermal Runaway 112
Thermocouple meter 141
Thermocouple Probes 204
Thermoelectric Systems 203
Tied Gyro 171
Torque Pressure Indicators 193
Torque/armature current characteristic 80, 81
Transfer Characteristic 246
Transformer Rectifier Unit 123
Transformer-coupled Two-stage Amplifier 271
Transformers 120
Transistor Amplifiers 262
Transistor Biasing 248
Transistor Circuit Configurations 250
Transistor Currents 249
Transistor Static Characteristics 255
Transistor Voltage Regulators 100
Transmitters 311
Transmitting Equipment 211
Triode as an amplifier 247
Triode Coefficients 246
Triode-Physical Characteristics 245
Triodes 245
Tropospheric Scatter Propagation 307
True Airspeed Indicator 165

Tunnel Diode 237
Tunnel Diode Oscillator 239
Tunneling Effect 236
Tunneling Theory 237
Turbine Exhaust Gas Temperature Indicators 204
Turn and Slip Indicators 173
Two-Cycle Errors 176
Two-layer Winding 56
Two's Complement Representation 281
Types of circuit connection 5
Types of Compass 174
Types of D.C. Motors 78
Types of Generators 61
Types of Resistors 16
Typical Pump-Driven Vacuum System 168
Typical System Operation 170

**U**

Units of Capacitance 189
Units of resistivity 15
Universal Ammeter Shunt 132
Use of the Ohmmeter 137
Uses of Lap and Wave Windings 57

**V**

V/I Characteristic 237
V/I Characteristics 229
Vacuum System 167
Vapour-pressure thermometer 201
Varactor DIODE 240
Variable Resistance systems 201
Various Gains of a CB Amplifier 263
Various Gains of a CC Amplifier 266
Various Gains of a CE Amplifier 264
Varistor (nonlinear resistor) 17
Varmeters 141
Vent Caps 111
Venturi-Tube Systems 167
Vertical Speed Indicators 165
VHF (Very High Frequency) Communications 313
Vibrating Contact Regulator 95
Vibrating-Reed Frequency Meter 143
Vibrator-type Voltage Regulator-SH-I 96
VLF Propagation 301
Voltage Buildup in Self-excited Generators 68
Voltage regulation 94
Voltage Coil Circuit Resistance 96
Voltage commutation 66
Voltage equation 79
Voltage equations 79
Voltage Recovery Check 114
Voltage Regulation 94
Voltage Regulator 95
Voltmeter 133
Voltmeter Accuracy 135
Voltmeter Sensitivity 135

**W**

Wattmeter 142
Wheatstone Bridge Circuit 202
Wire-Wound 17
Working 229, 231, 233
Workstations 293
Wound rotor or slip ring rotor 85

**Y**

Yoke 52

**Z**

Zener Diode 100
Zener Diode 234
Zener Diode as Peak Clipper 235
Zener Diode As Voltage Regulator 235
Zener Diode For Meter protection 235