15BEEC7OE02 **CONSUMER ELECTRONICS** **3 0 0 3 100**

OBJECTIVES

- To study about various speakers and microphone
- To learn the fundamental of television systems and standards
- To learn the process of audio recording and reproduction
- To study the various telephone networks

INTENDED OUTCOMES:

- Gain knowledge about various speakers and microphone
- Gain knowledge about the fundamental of television systems and standards
- Gain knowledge about the process of audio recording and reproduction
- Gain knowledge about the various telephone networks

UNIT I LOUDSPEAKERS AND MICROPHONES

Dynamic Loudspeaker, Electrostatic loudspeaker, Permanent Magnet Loudspeaker, Woofers and Tweeters - Microphone Characteristics, Carbon Microphones, Dynamic Microphones and Wireless Microphones.

UNIT – II TELEVISION STANDARDS AND SYSTEMS

Components of a TV system – interlacing – composite video signal. Colour TV – Luminance and Chrominance signal; Monochrome and Colour Picture Tubes - Colour TV systems – NTSC, PAL, SECAM - Components of a Remote Control.

UNIT – III OPTICAL RECORDING AND REPRODUCTION

Audio Disc – Processing of the Audio signal –read out from the Disc – Reconstruction of the audio signal – Video Disc – Video disc formats- recording systems – Playback Systems.

UNIT – IV TELECOMMUNICATION SYSTEMS

Telephone services - telephone networks – switching system principles – PAPX switching – Circuit, packet and message switching, LAN, MAN and WAN, Integrated Services Digital Network. Wireless Local Loop. VHF/UHF radio systems, Limited range Cordless Phones; cellular modems

UNIT – V HOME APPLIANCES

Basic principle and block diagram of microwave oven; washing machine hardware and software; components of air conditioning and refrigeration systems.

Text Book:

1.S.P.Bali, "Consumer Electronics", Pearson Education, 2005.

Reference Book:

1. R.G.Gupta,"Audio & Video Systems",Tata Mc Graw hill Publishing Company Ltd, 2004.
2.Ajay sharma,"Audio Video & TV Engineering consumer electronics ", Dhanpat Rai & Co.(P).Ltd,2007

STAFF IN CHARGE **HOD**

KARPAGAM ACADEMY OF HIGHER EDUCATION
(Deemed to be University Established Under Section 3 of UGC Act 1956)
Pollachi Main Road, Eachanari Post,
Coimbatore – 641 021
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

LECTURE PLAN

NAME OF THE STAFF : Mrs.P.SASIKALA

DESIGNATION : ASSISTANT PROFESSOR

CLASS : B.E-IV YEAR EEE & CSE

SUBJECT : CONSUMER ELECTRONICS

SUBJECT CODE : 15BEEC7OE02

S.No	TOPICS TO BE COVERED	TIME DURATION	SUPPORTING MATERIALS	TEACHING AIDS
UNIT I LOUDSPEAKERS AND MICROPHONES				
1	Introduction to consumer Electronics	01	R1- Page.no 50	BB
2	Dynamic Loudspeaker	01	R1- Page.no 50	BB
3	Electrostatic loudspeaker	01	R1 Page.no 41-50	BB
4	Permanent Magnet Loudspeaker	01	R1- Page.no.52	BB
5	Woofers and Tweeters	01	R1 Page.no.61-64	PPT
6	Microphone Characteristics	01	R1 page.no.18	PPT
7	Carbon Microphones	01	R1 page.no.36	PPT
8	Dynamic Microphones	01	R1 page.no.33-41	PPT
9	Wireless Microphones.	01	R1 page.no.40	PPT
Introduction		**01**		
Total Lecture Hours		**08**		
Total Hours		**09**		

UNIT – II TELEVISION STANDARDS AND SYSTEMS				
10	Components of a TV system	01	R1 Page.no 304	Smart Board
11	Interlacing Scanning	01	R1 Page.no 308	BB, PPT
12	composite video signal.	01	R1 Page.no 311	PPT
13	Colour TV – Luminance and Chrominance signal;	01	R1 Page.no 371	PPT
14	Monochrome and Colour Picture Tubes	01	R1 Page.no 342-353	PPT
15	Colour TV systems – NTSC	01	R1 Page.no 375	PPT
16	PAL	01	R1 Page.no 354	PPT
17	SECAM	01	R1 Page.no 388	PPT
18	Components of a Remote Control.	01	www.television.remote.com	PPT
Total Lecture Hours		**09**		
Total Hours		**09**		

UNIT – III OPTICAL RECORDING AND REPRODUCTION				
19	Audio Disc	01	R1 Page.no.290	Smart board
20	Processing of the Audio signal	01	R1 Page.no 290	Smart board
21	read out from the Disc	01	R1 Page.no 292	Smart board
22	Reconstruction of the audio signal	01	R1 Page.no 301	Smart board
23	Video Disc	01	R1 Page.no 262	Smart board
24	Video disc formats	01	R1 Page.no 264	Smart board
25	Recording systems	01	R1 Page.no 275 & 286	Smart board
26	Playback Systems	01	R1 Page.no.286	Smart board
27	Revision of Recording Systems	01	-	Smart board
Total Lecture Hours		**09**		
Total Hours		**09**		

UNIT – IV TELECOMMUNICATION SYSTEMS				
28	Telephone services	01	R1 Page.no.1-42	BB
29	telephone networks	01	R1 Page.no.1-5	BB
30	switching system principles	01	R1 Page.no.16	BB
31	PAPX switching Circuit	01	R1 Page.no.348	BB
32	packet and message switching,	01	R1 Page.no.414	BB
33	LAN, MAN and WAN,	01	R1 Page.no.455-467	BB
34	Integrated Services Digital Network.	01	R1 Page.no.492-559	BB
35	Wireless Local Loop. VHF/UHF radio systems,	01	R1 Page.no 560-586	BB
36	Limited range Cordless Phones;, cellular modems	01	R1 Page.no 560-586	BB
Total Lecture Hours		**09**		
Total Hours		**09**		

UNIT – V HOME APPLIANCES				
37	Basic principle and block diagram of microwave oven;	01	R1 Page.no 487	PPT
38	Basic principle and block diagram of microwave oven;	01	R1 Page.no 487	PPT
39	washing machine hardware and software;	01	T1. www. safaribooksonline.com	PPT
40	washing machine hardware and software;	01	T1. www. safaribooksonline.com	PPT
41	components of air conditioning	01	www.aceac.com	PPT
42	components of air conditioning	01	www.aceac.com	PPT
43	Refrigeration Systems	01	//berg.group.com/	PPT
44	Refrigeration Systems	01	//berg.group.com/	PPT
45	Revision	01	-	PPT
Total Lecture Hours		**09**		
Total Hours		**09**		

Total No of Hours for Introduction: 01 Hrs

Total No of Lecture Hours Planned: 44 Hrs

Total No of Hours Planned : 45 Hours

TEXT BOOKS:

S.NO.	Author(s) Name	Title of the book	Publisher	Year of the publication
1.	S.P.Bali	Consumer Electronics	Pearson Education	2005

REFERENCE BOOKS:

S.NO.	Author(s) Name	Title of the book	Publisher	Year of the publication
1.	R.G.Gupta	Audio & Video Systems	Tata Mc Graw hill Publishing Company Ltd	2004
2.	Ajay sharma,	Audio Video & TV Engineering consumer electronics	Dhanpat Rai & Co.(P).Ltd	2007
3.	Manav Bhathagar	Telecommunication Switching systems & Network	PHI Learning Pvt.Ltd	2015

STAFF IN-CHARGE **HOD/ECE**

UNIT I

LOUDSPEAKERS AND MICROPHONES

All *sound recording* starts with the use of microphones. Professional recording engineers use microphones in large number, with the output from each microphone being *separately* recorded on a wide strip of recording tape. For example, 32 or more microphones are used to record a large orchestra. The *placing* of each microphone, the *amplitude* of recording from each microphone, and the subsequent *mixing* of the sound from each track to form two tracks of a *stereo recording* or four tracks of a discrete *quad recording* are operations which require great skill and experience.

A microphone is a device of the class called *transducers* which converts sound waves in air into electrical waves of the same frequency and shape. In the process of conversion, the microphone must make use of either the *pressure* of the air waves, or the *velocity* at which the air moves. So, we have two types of microphones, the pressure-operated types, and the velocity-operated types.

CHARACTERISTICS OF MICROPHONES

There are many types of microphones available. Each has certain advantages and disadvantages. The choice of a microphone depends upon the type of material to be reproduced, the placement of the microphone, whether it is to be used indoors or outdoors, the frequency response desired, and a number of other factors.

The basic types of microphones, grouped according to their *principle of operation* are : carbon, crystal, dynamic, ribbon and capacitor.

Each of these has its own characteristics with respect to : (1) output level, (2) frequency response, (3) output impedance and (4) directivity. These characteristics ultimately determine the particular type of microphone suitable for a given application.

Output Level

The output level of a microphone governs the amount of amplification that must be available for use with the microphone. The output level of microphones is usually given in dB preceded by a minus sign. *The minus sign means that the output level is so many dB below the reference level of 1 milliwatt for a specified sound pressure.*

The unit of sound pressure used for rating microphones is referred to as a *bar*. A bar is equal to a sound pressure of 1 dyne per square centimetre. *Speech* provides sound pressures between 0.4 and 15 bars. For *music* the pressure ranges from 0.5 bars to 1250 bars.

Microphones are rated in a number of different ways, and this often causes confusion. If ratings are given in any manner other than in bars, it is a good idea to *convert* their output level rating to dB below 1 milliwatt for a sound pressure of 1 bar (see Table 2.1).

Table 2.1 **Comparison of Microphone Ratings**

Rating Given	Correction Factor
dB below 1 m W/1 bar	0 dB
dB below 1 mW/10 bars	– 20 dB
dB below 1 volt/1 bar	2 dB
dB below 1 volt/10 bars	– 18 dB

A microphone with a *low output level* necessitates the use of an *amplifier with greater gain*, which, in turn increases the possibility of noise and hum.

Frequency Response

The frequency response of a microphone is a rating of the fidelity of relative output voltage which results from sound waves of different frequencies. The simplest way to find a complete picture of the frequency response characteristics of a microphone is to plot a curve of its output voltage vs input frequency. *Since good modern microphones are relatively flat over their range, it is often considered sufficient to specify the range over which their output does not vary more than plus or minus 1 or 2 dB.*

For ordinary home high-fidelity use, a microphone *frequency-response curve* should be reasonably flat between 40 and 10,000 Hz. With systems designed specifically for speech reinforcement, a **lower limit** of 150 Hz and an **upper limit** of 5,000 Hz is entirely satisfactory. Where it is desired to reproduce music with the highest possible fidelity, the frequency response should be flat (within 2 dB) from about 40 to 15,000 Hz, Fig. 2.1 shows the response of several types of microphones.

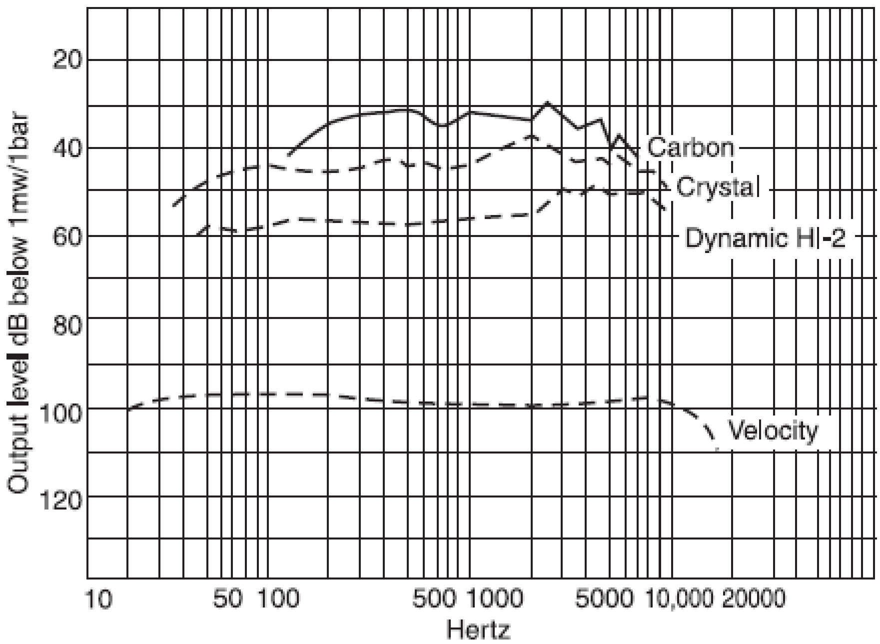


Fig. 2.1 Frequency response curves for typical microphones

Output Impedance

A microphone, like any other component with electrical inputs or outputs, has a value of impedance. When a microphone is connected to an amplifier, a *complete circuit* is formed and electric current flows whenever a sound causes the microphone to **generate** an electrical voltage.

For most high quality microphones impedance is *low*, a few ohms ranging up to a hundred ohms or so, but as little as a fraction of an ohm in a ribbon microphone. Only capacitor and ceramic crystal microphones have *high* impedances.

The importance of microphone impedance is not a matter of the precise value but of *the ability of the microphone and the recorder to be matched together*. High impedance microphones must be connected into a recorder with high impedance input, otherwise both the signal amplitude and the frequency range will be adversely affected.

In general, when a ceramic or capacitor microphone is connected to a low impedance amplifier input, the output from the amplifier is small and the lower frequencies are reduced disproportionately, producing an unpleasant shrill sound on playback. Cassette recorders of good quality are not as a rule designed to take *high impedance microphones*, except when such microphones are supplied with the recorder, and if, for any reason, such microphones are used, a high impedance preamplifier must be used between the microphones and the recorder.

Low impedance microphones and inputs are much more common; one good reason is the *hum problem*. The frequency of the supply voltage, 50 Hz, is, like any other ac signal, radiated from any wire on which it is present, and can also be picked up on any wire. You can prove this for yourself by touching the input of an amplifier with the volume control turned half-way up. The resulting amplified hum sounds very loud in the loudspeakers, showing that the *hum signal* must be at least as strong as the signals normally fed into the amplifier. *Radiation* through the air at this low frequency is not an efficient process, however, as the house wiring is not long enough to act as an efficient antenna (aerial), and the hum signal picked up on a piece of wire behaves as if it had come from a high impedance source. If the input of the cassette recorder amplifier is at high impedance, then the whole hum signal can be *transferred* to it unless every piece of wire and every connector is efficiently *shielded* from the hum by *earthed metal casings.* If, on the other hand, the input to the cassette recorder amplifier is at low impedance, the hum signal, if picked up, is *greatly reduced in amplitude* even if no screening is used.

Directivity

Microphones do not respond equally to sound reaching them from all directions. Their *frequency response characteristics* also vary, depen+ding on the *angle* at which the sound reaches them. A microphone may respond *equally* to all frequencies between 40 and 10,000 Hz when the sound is originating directly in front of it, while the high-frequency response falls off rapidly as the sound originates further to *either side.* Where it is necessary to pick up sound from all directions, the directional characteristics of all microphones are not suitable.

The way in which a microphone responds to sounds coming from different directions is plotted on a circular graph which is known as a *polar diagram.* The centre of the circle is the zero point and concentric circles indicate successively higher levels of response as they move outward. The top of the circle is the *front* and the bottom the *back* of the microphones, and the straight lines radiating from the centre denote the *corresponding angles.*

A true *omnidirectional* microphone would have a plot of a perfect circle, Fig. 2.2, hence there would be little point in making a diagram. Few manufacturers in fact publish polar diagrams of such instruments. Actually, the pickup is rather less at the *rear* than at the *front*, so the plot would have a smaller radius at the bottom than at the front. *An omnidirectional microphone is most commonly used for general recording, and is the best for recording sound sources that move about a lot.*

When the directional response of a *pressure gradient* microphone is plotted, the shape of the figure is something like a heart, with the *maximum* response in the front, curving around the sides to a *minimum* near the rear. Therefore, it is termed *cardioid* (Greek, kardia = heart), Fig. 2.3.

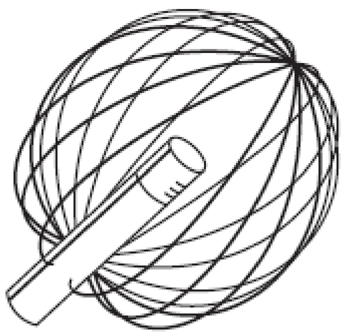


Fig 2.2 Omnidirectional pickup pattern

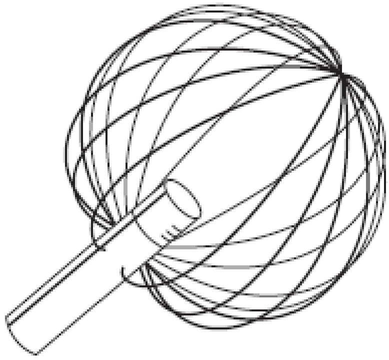


Fig. 2.3 Cardioid pickup pattern

To achieve a greater rejection of sounds coming from the sides, the *vents* in some models are designed to admit a higher sound pressure to give more cancellation at the rear of the cone or diaphragm. The result can be seen in the polar diagram, Fig. 2.4, as a narrower forward lobe, and the response is known as *hyper* or *supercardioid.* One effect of this is that *rear sounds* can actually exert more pressure on the back of the cone than on the front, so movement is produced, although it is negative, that is, a high pressure region produces forward instead of backward movement of the cone.

This appears on the diagram as a small lobe sometimes marked with a *negative sign* at the rear or below the centre position. It means that there is some pickup at the rear although not so much as at the front, but *the point of minimum pickup is nearly always about 120° from the front.* The rear lobe varies considerably with frequency and is sometimes found with cardioid as well as the hypercardioid.

If both sides of the ribbon of a velocity microphone are open to free air, it follows that sounds from front or back will have exactly the same effect. In fact, *there is no front or back.* As the ribbon is actuated by **particle velocity** instead of pressure, it responds only when *facing* the sound source. Sounds coming from the sides exert equal force on both sides of the ribbon and so have no effect. It is, therefore, *bidirectional*, and its diagram consists of two near circular lobes looking like a figure 8, hence the designation of a *figure of eight response*, Fig. 2.5.

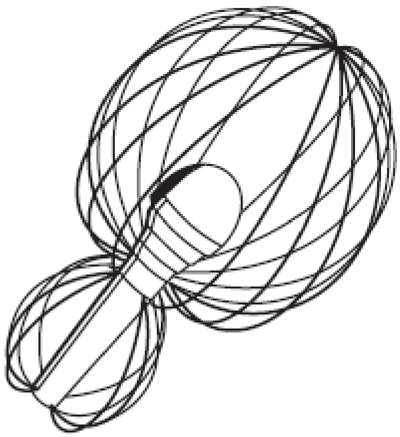


Fig. 2.4 Hyper or super-cardioid pickup pattern

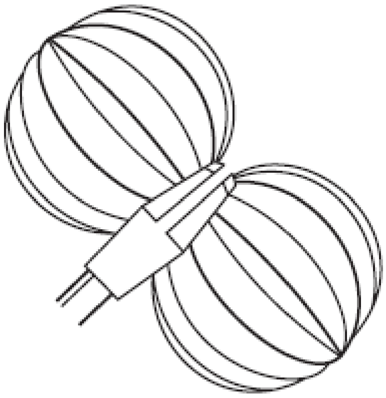


Fig. 2.5 Figure-of-eight pickup pattern

Example 2.1 : The voltage gain of an amplifier, when it feeds a resistive load of 1.0 kΩ, is 40 dB. Determine the magnitude of the output signal voltage and the signal power in the load when the input signal is 10 mV.

Solution

$$A_P = 20 \log \frac{V_2}{V_1} = 40$$

$$\log \frac{V_2}{V_1} = 2.0$$

$$\frac{V_2}{V_1} = 100$$

$$V_2 = 100 \times 10 = 1000 \text{ mV} = \mathbf{1.0\ V}$$

$$P_2 = \frac{V_2^2}{R_L} = \frac{1.0^2}{1000} = 0.001\ \mathbf{W = 1\ mW}$$

Example 2.2 Express the power dissipated in a 15 Ω resistor in decibel relative to 1 mW when the voltage across the resistor is 1.5 V rms.

Solution

$$P_2 = \frac{V^2}{R} = \frac{1.5^2}{15} = 0.15 \text{ W} = 150 \text{ mW}$$

$$A_p = \frac{P_2}{P_1} = \frac{150}{1} = 150$$

and $A_p = 10 \log 150 = 10 \times 2.176$

$= \mathbf{21.76\ dB}$

Example 2.3 An audio amplifier produces 20 watt output across an 8 ohm resistance when a 5 millivolt signal is applied to its input across a 1 mega ohm resistor. Determine the decibel gain.

Solution

The output power, $P_1 = 20$ watt

Input power, $P_2 = \frac{E_2^2}{R_2} = \frac{(0.005 \text{ volt})^2}{10^6 \text{ ohm}} = \frac{25 \times 10^{-6}}{10^6} = 2.5 \times 10^{-11}$ watt

Hence, $\text{dB} = 10 \log \frac{P_1}{P_2} = 10 \log \frac{20 \text{ watt}}{2.5 \times 10^{-11} \text{ watt}}$

$= 10 \log 8 \times 10^{11} = 10 \times 11.903 = \mathbf{119.03\ dB}$

Alternatively, the output voltage

$$E_1 = \sqrt{P_1 \times R_1} = \sqrt{20 \text{ watt} \times 8 \text{ ohm}}$$

$$= \sqrt{160} \text{ volt} = 12.65 \text{ volt}$$

$$\text{dB gain} = 20 \log \frac{E_1}{E_2} + 10 \log \frac{R_2}{R_1}$$

$$= 20 \log \frac{12.65 \text{ volt}}{0.005 \text{ volt}} + 10 \log \frac{10^6 \text{ ohm}}{8 \text{ ohm}}$$

$$= 20 \log 2530 + 10 \log 1.25 \times 10^5$$

$$= 20 \times 3.403 + 10 \times 5.0969$$

$= 68.06 + 50.97 = \mathbf{119.03\ dB}$ (as above)

If the resistance ratio $\left(\frac{R_2}{R_1}\right)$ had been ignored, an *erroneous gain figure* of about 68 dB would have been obtained.

Example 2.4 A microphone has an output of – 60 dB (with respect to a zero level of 6 milliwatt) and is connected to the 0.5 mega ohm input of a preamplifier. The preamplifier has a gain of + 40 dB. The signal then passes through an equaliser with an insertion loss of – 15 dB and through a main amplifier with a gain of + 65 dB. If the output to the speaker is 6 watt, find the total power gain and the input voltage to the preamplifier.

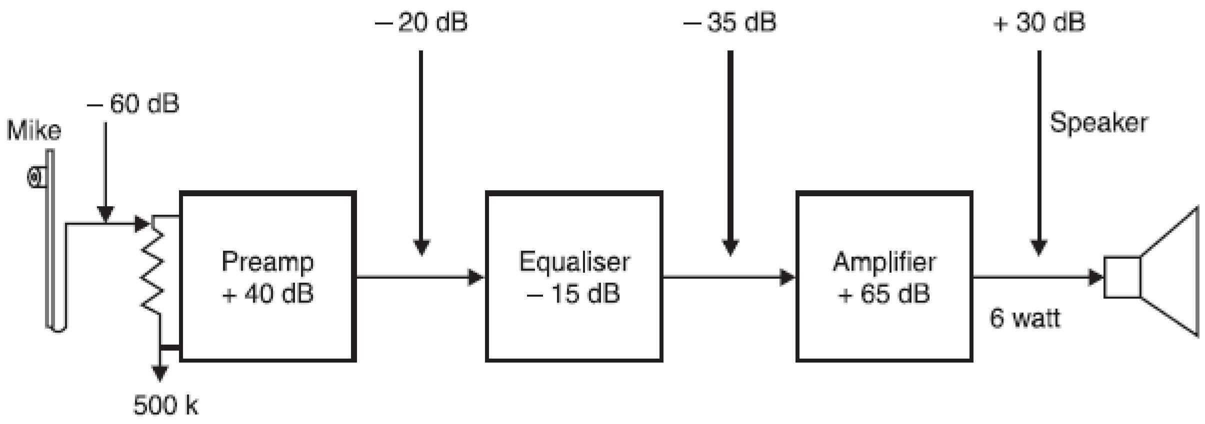


Illustration for Example 2.4

Solution *Decibel gains and losses can be added algebraically. Hence,*

Total dB gain = + 40 dB – 15 dB + 65 dB **= + 90 dB**

The output to the speaker, therefore, is

– 60 dB (input) + 90 dB (gain) = + 30 dB

Since $90 \text{ dB} = 10 \log \frac{P_1}{P_2}$, $9 = \log \frac{P_1}{P_2}$ and $\frac{P_1}{P_2} = \log^{-1} 9 = 10^9$

Output power $P_1 = 6$ watt;

hence, $P_2 = \frac{P_1}{10^9} = \frac{6 \text{ watt}}{10^9}$

$= 6 \times 10^{-9}$ watt

Input voltage $E_2 = \sqrt{P_2 R_2} = \sqrt{(6\times10^{-9})\times(0.5\times10^6)}$

$= \sqrt{0.003}$ volt

$= 0.0548$ volt $=$ **54.8 mV**

CARBON MICROPHONES

For a *telephone system* important requirements are *(i)* the microphone shall be of convenient size; *(ii)* capable of mass production at low cost while possessing high sensitivity to operate from a simple battery; *(iii)* its performance must be stable and adequate to provide intelligible speech and articulation and; *(iv)* it need not necessarily include the higher harmonic frequencies for reproducing.

On the other hand, microphones used for purposes such as *radio broadcasting* are relatively few in number and their cost is not a primary consideration: high fidelity reproduction up to about 10,000 Hz for natural speech and music transmission is essential.

An inset pattern of *carbon granule microphone* is in general use for telephone systems: the usual type is a *self-contained and sealed* microphone which can be readily and completely removed from the telephone instrument. Its operation depends upon the variation in *contact resistance* of the carbon granules when they are subjected to the **pressure changes** of sound waves. It follows that *this type of microphone does not produce an emf but functions by modulating the current obtained from an external battery.*

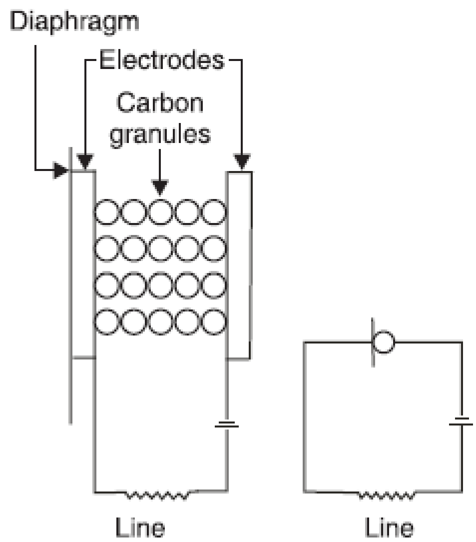


Fig. 2.6 Principle of carbon–granule microphone

The essential components of this microphone are two electrodes and carbon granules which are loosely packed in between the electrodes. One electrode is *fixed* relative to the other which carries a *diaphragm* to respond to the pressure changes of the sound waves. *Movements of this diaphragm vary the resistance of the granules and so control the line current in accordance with the sound waves reaching the diaphragm.* An increase in pressure produces a reduction in resistance and an increase in current. The elementary circuit arrangement is shown in Fig. 2.6. Constructional details of the inset carbon microphone are shown in Fig. 2.7.

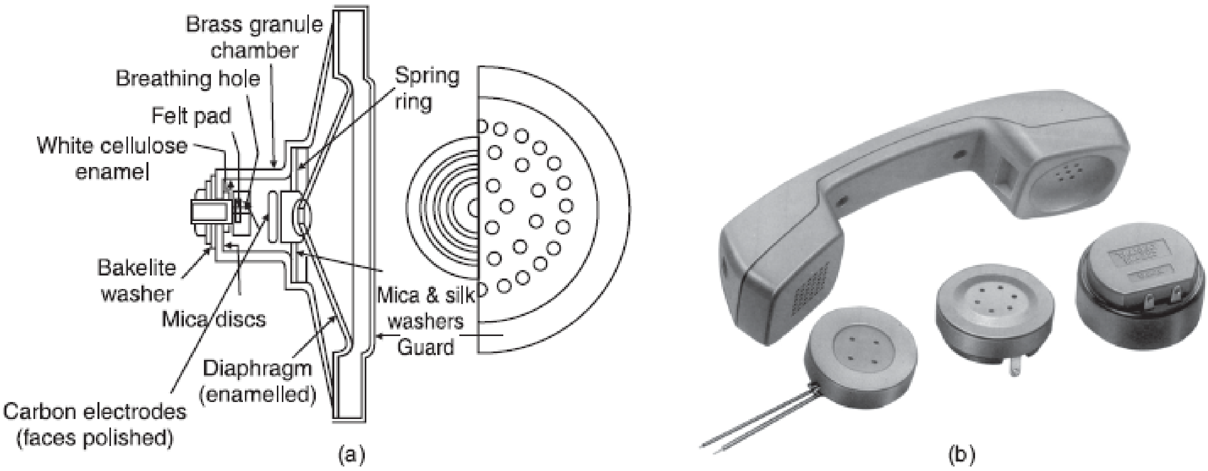


Fig. 2.7 Inset type carbon granule microphone (a) Constructional details, (b) Carbon microphone in a telephone handset

The carbon microphone generates a *continuous hiss.* This hiss is due to small variations in contact resistance which take place between the carbon granules.

With carbon microphones the electrical output is not directly proportional to the sound input level. The practical effect of this non-linear distortion is to produce *harmonics* of the lower speech frequencies and these harmonics tend to *mask* higher frequencies normally present in the speech, resulting in *loss of clarity or articulation.*

The average *output level* of carbon microphone is of the order of –30 dB (see Fig. 2.1). The best carbon microphones have a *frequency response* of approximately 60 to 7,000 Hz. They are substantially **non-directional** although their high frequency response above 300 Hz usually falls off at angles exceeding 40 degrees from the front of the microphone.

When the maximum output level is required from a microphone, the carbon microphone is often used. The frequency response characteristics of the carbon microphone are poor and cannot be used for high-fidelity work.

CRYSTAL MICROPHONES

Certain crystals, such as rochelle salt and quartz possess the property of *generating small emf's when subject to stress or strain.* This effect is utilised in what is known as the *crystal microphone.*

The construction of a crystal microphone is shown in Fig. 2.8. A thin finger shaped slice of crystal is secured at one end by means of a compliant clamp, and the apex of a cone is made to bear against the other. Sound pressure waves cause the cone to alternately press against and bend the crystal slice and release it. Thus, corresponding voltages are generated across the slice. A pair of contacts is fixed to opposite surfaces to take off the signal.

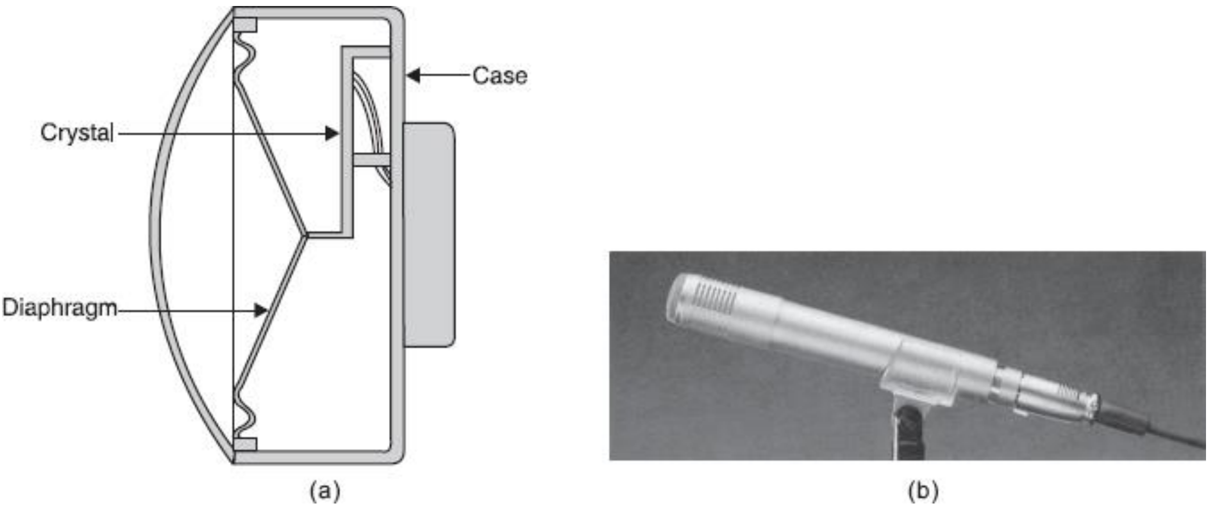


Fig. 2.8 Simple crystal transducer: (a) Crystal slice is secured at one end with compliant clamps and bending action caused by pressure from the cone generates an emf across the slice. (b) A typical crystal microphone

An improvement is obtained if the single slice of crystal is replaced by *two slices cemented together*. Then, when pressure is exerted, one slice is *compressed* while the other is *stretched*. Thus *equal and opposite* voltages are produced which, being *in series* like the cells of a car battery, give *double* the output. Any non-linearity which may arise due to the different mechanical strains between pressure and release is also thereby compensated. The double crystal unit is termed as *bimorph.*

One of the snags with this type of transducer is *the mass which must be moved* by the sound pressure acting on the cone. This consists of the mass of the cone plus that of the crystal, or that part of it which is moved. This restricts the frequency response to its upper end to around 10 kHz and also limits the transient response. In addition there are **resonances** due to the cone and the crystal.

With some of the better microphones the cone does not actuate the crystal directly but through a *cantilever*. According to the dimensions involved, the effect of the mass of the crystal and the mechanical resistance offered by its stiffness can be reduced but at the same time so also is the amplitude of the transmitted vibrations, hence the signal output.

Another type of construction is the *sound cell* where several crystal elements are sealed together, this also being termed as *multimorph.* Here the cone is often dispensed with, the sound pressure waves acting *directly* on the crystal. Output is lower with this arrangement, but the frequency response is better and also the cone resonance is eliminated.

There is no dc path through a crystal microphone, the crystal being an insulator. Having the two electrical contacts on either side of the slice, the unit behaves as a capacitor. The *equivalent circuit*, then, consists of a voltage source in series with a capacitor, Fig. 2.9. Capacitance values vary, but around 1,000 pF (0.00μF) is typical. This should be taken into account when considering *cable requirements.*

The crystal microphone is the type most widely used in lower cost installations. It has a relatively *high output level* and a *high impedance.* A long cable will reduce the output voltage available from a crystal microphone and may affect its high frequency response.

The *output level* of this type of microphone is usually between –48 dB and –60 dB. Their output impedance is almost always more than 100,000 ohms.

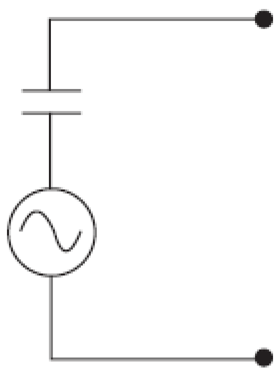


Fig. 2.9 Equivalent circuit of a crystal microphone consisting of an ac source in series with a capacitor

Good units may have a *frequency response* substantially flat between 50 and 10,000 Hz. Units are also available with slightly wider frequency response ranges.

The crystal microphone is normally non-directional although a pressure-gradient crystal microphone which gives a unidirectional response pattern is also being marketed. This microphone gives excellent results.

The natural crystals, such as rochelle salt, are not very durable. They are adversely affected by humidity and high temperature; also they fracture easily when subjected to shock. If a crystal microphone is subjected to a temperature of 130 degrees, it will be rendered completely useless. *Care must always be taken to avoid exposing a crystal microphone to direct sunlight for any length of time.*

MOVING COIL (DYNAMIC) MICROPHONES

We now come to one of the *most common* type of transducers, the principle being used both for microphones and loudspeakers.

This type of microphone [Fig. 2.10(a)] works on the *generator principle.* The diaphragm carries a coil of wire placed in an intense magnetic field of *constant value* (B). Movements of the diaphragm, consequent upon sound pressure changes, result in an emf ($E \propto Blv$) being generated in the conductor. This emf is proportional to the velocity of motion (v) of the conductor in the air-gap: if the microphone is to have the same sensitivity at all frequencies it is necessary for the velocity of motion of the coil, due to a sound of given intensity, to be independent of the pitch of the sound.

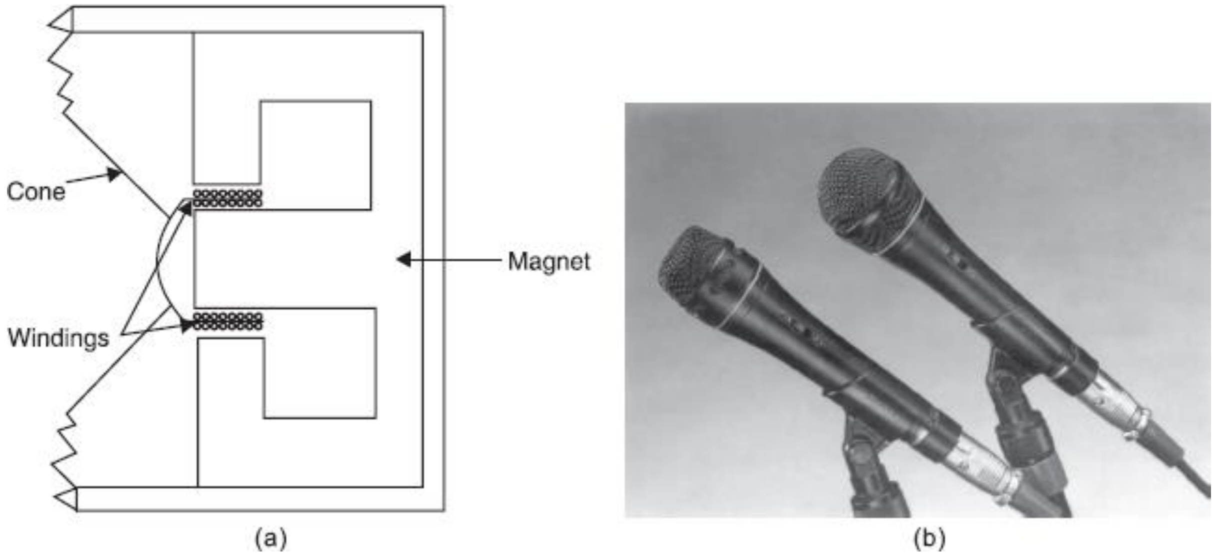


Fig. 2.10 Moving–coil microphone (b) Different models of dynamic microphones

The ends of the coil are secured to two points on the cone and from these a pair of loose flexible wires is connected to terminal tags to which the microphone cable may be connected. So the electrical replica or *signal* as it is termed, can be conveyed to an external circuit.

In order to respond to transients and high frequency sound waves it is necessary to keep the mass of the cone and its coil as small as possible. For this reason, *the coil is usually wound with aluminium wire of very fine gauge and the number of turns is limited.* This means that the generated voltage is also limited, and so *the natural impedance of the instrument is low*. The average is around 30 Ω, but many models have a transformer incorporated in the instrument itself to give an output at a higher impedance.

Even with these measures, the combined mass of the cone, which must be rigid to faithfully respond to the pressure waves and also keep the coil accurately *centred* within the pole-pieces, restricts *transient response* and more seriously places the *mechanical resonance* well within the frequency range of the instrument.

The moving-coil unit is reliable and robust, features which fit it for applications calling for hard usage. It can be made quite cheaply and therefore fills the need for an *inexpensive* *microphone*, but refinements and careful design can add considerably to the cost, and so the better units are quite expensive.

The *output level* of most dynamic microphones is about 55 or more dB below 1 milliwatt per bar. The ordinary dynamic microphone is essentially nondirectional, although its high-frequency response falls off rapidly on either side as shown in Fig. 2.11. To make full use of a dynamic microphone's frequency range, the microphone should face directly toward the source of sound.

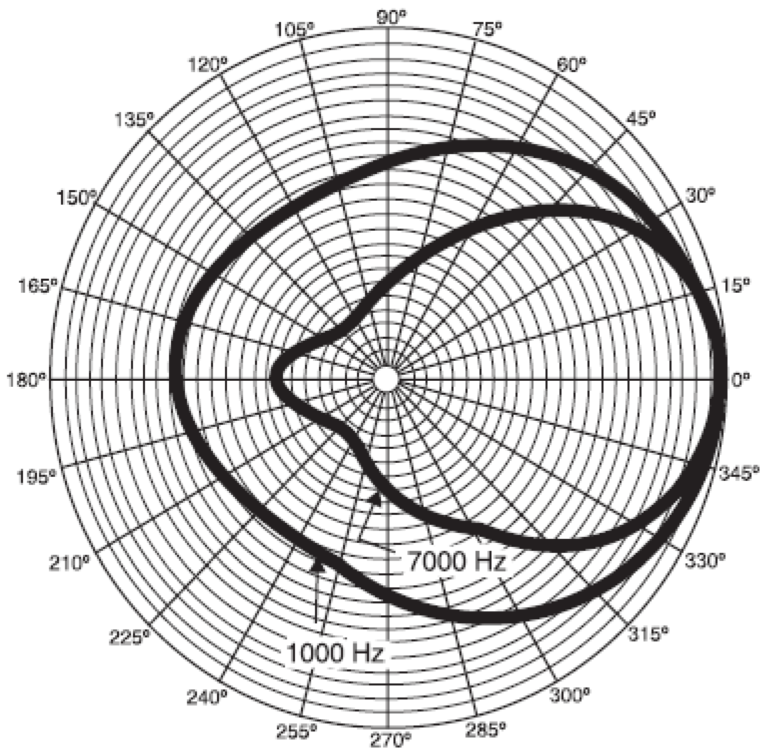


Fig. 2.11 Directivity of a dynamic microphone

One of the refinements, sometimes found in the better models, is what is known as a *hum bucking coil*. An ac operated equipment, especially that containing a mains transformer, is often surrounded by a magnetic field which oscillates at the mains supply frequency (50 Hz in India). If a moving coil microphone were used with such a field, a voltage of 50 Hz may be induced into the coil. Although this voltage would be quite small so also is that produced by cone movement, hence it could be a significant level as compared to the wanted signal. After amplification the result would be an *audible hum*. The hum bucking coil is wound in the *opposite direction* to the moving coil and is positioned close to the moving coil capsule. Any hum field affects both the coils but as they are *connected in series and in opposite phase,* it is cancelled by the production of equal and opposite emf.

Moving-coil microphones are often described as *dynamic* but the term is sometimes applied to all transducers that operate on the electromagnetic principle. Another type is the ribbon unit.

RIBBON (VELOCITY) MICROPHONES

With these microphones, *the sound operates the transducing element directly without the use of a cone*. A ribbon of aluminium foil, *corrugated to allow backward and forward motion*, is mounted edgeways between two pole-pieces [Fig. 2.12(a)]. We have here, then, the classic example of electromagnetic generation, a single conductor moving in a powerful magnetic field. There is no multiplication of induced emfs by the successive turns of a coil, *the ribbon can be regarded as a coil with but a single turn.* As a result, the output voltage is very low and so also is the impedance, something of the order of 0.1Ω. *All ribbon microphones, therefore, have a built-in transformer to step-up the impedance and voltage to a usable level.*

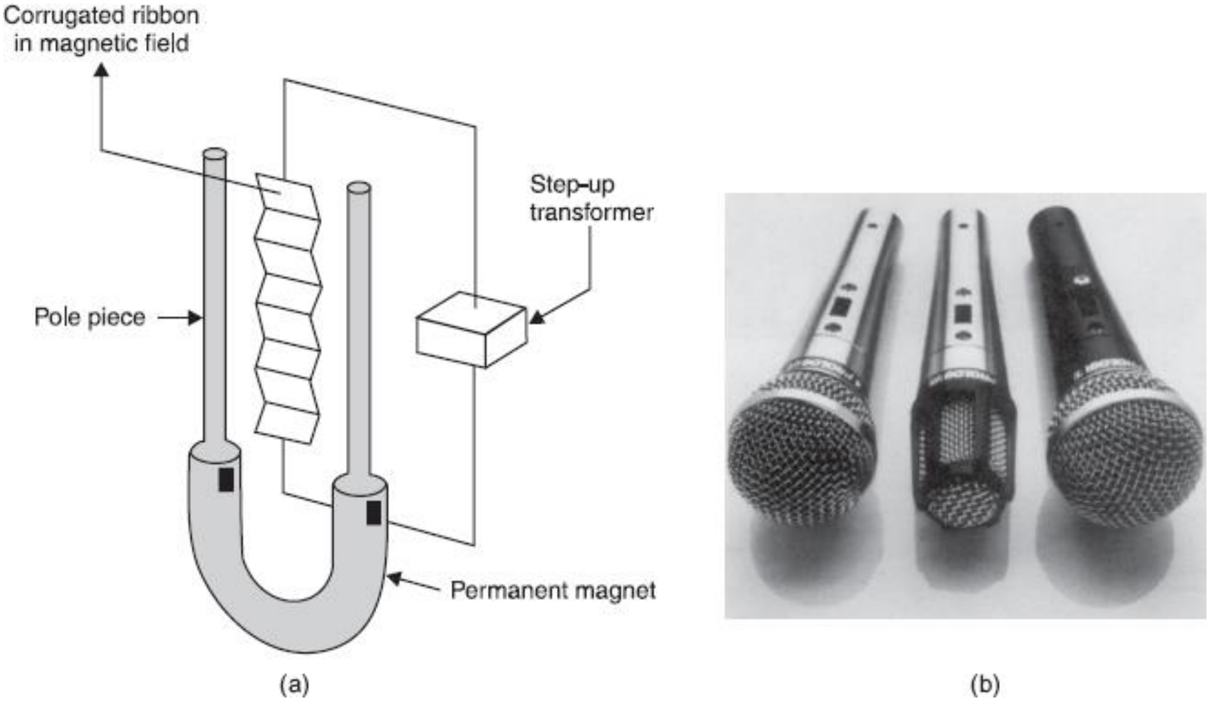


Fig. 2.12 (a) Ribbon microphone consists of a single conductor (the ribbon) in a magnetic field and so always includes a transformer to step up voltage and impedance. (b) Ribbon microphones Schure

Even with the transformer, the output voltage for any given impedance value is still lower than that for a moving coil unit. It depends on the number of magnetic lines of force cut by the ribbon, which in turn depends on the strength of the magnet and the length of the ribbon. To achieve sufficient output, older microphones used large magnets and long ribbons which not only made them bulky and heavy but introduced *ribbon resonances* well within the audible range of frequencies.

More recent developments have produced materials that enable magnets to be made *powerful yet light* and also *flared front apertures* that increase the acoustic force acting on the ribbon. This has enabled ribbon and magnet size to be reduced and units with an adequate output are now available with ribbon sizes less than 1 inch in length and with a mass of less than half a milligram. *This has enabled an even smoother and extended frequency response to be obtained with freedom from resonance*

The *output level* of a velocity microphone is usually 60 dB below 1 milliwatt per bar. Generally ribbon microphones have excellent response characteristics.

The ribbon microphone is bidirectional. Maximum response is to sound reaching the front or back of the microphone at a 90-degree angle to the plane of the ribbon faces. It is more directional than the crystal and dynamic microphones, the overall response of the ribbon microphone falls off as the angle of sound reaching it varies from 90 degrees to the faces of the ribbon (Fig. 2.13).

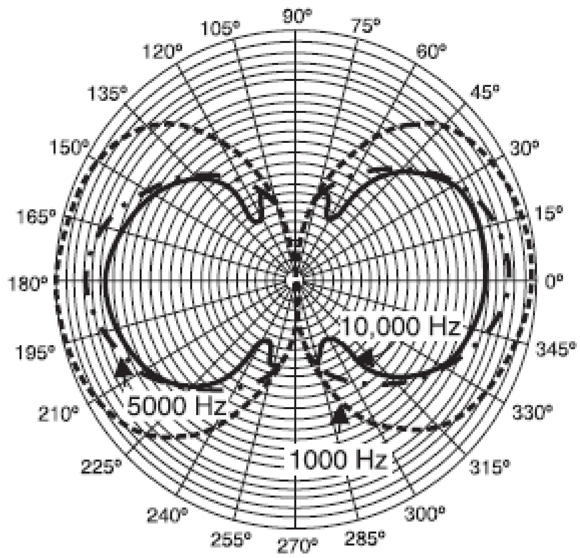


Fig. 2.13 Directivity of a ribbon microphone

The ribbon microphone is quite sensitive to the movement of the air surrounding it, and *it must be carefully protected from puffs of wind when used outdoors.* A ribbon microphone should be placed at least 18 inches from the source of the sound.

CAPACITOR MICROPHONES

A capacitor (or to give it its original name, *condenser)* microphone, Fig. 2.14, is one that depends for its operation on the *variation of capacitance* between a fixed plate and a tightly stretched metal diaphragm. Its development represents a milestone in the history of modern electroacoustics and for a number of years *this type of microphone was the accepted standard for high-quality sound systems.*

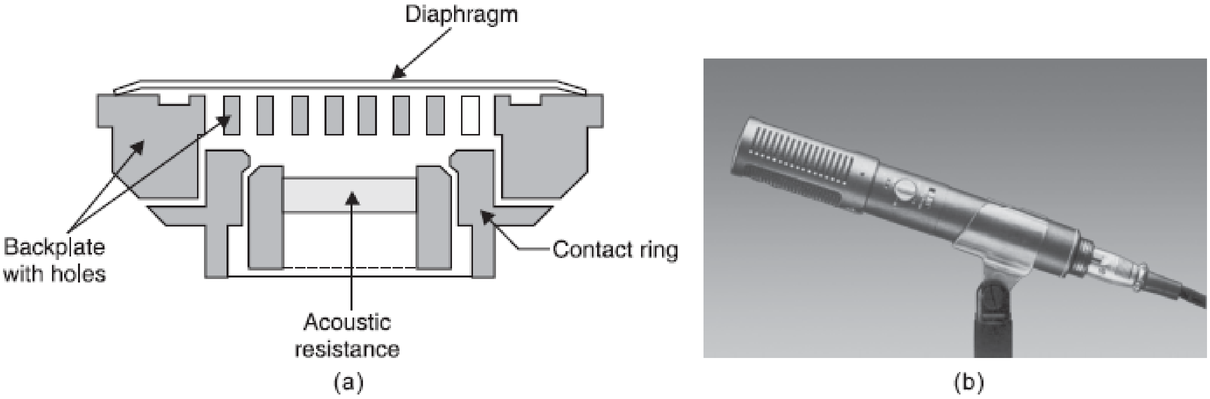


Fig. 2.14 (a) AKG capacitor microphone—cross-section (b) Capacitor microphone

In this case the movable plate is the diaphragm which serves as the *sound pressure sensing element.* In order that it will respond well to transient and high frequency sound waves, the diaphragm is made very *light* by forming it from a thin film of plastic material coated with a fine layer of metal. Aluminium is sometimes used for its *lightness*, but gold is also used for *anticorrosion properties* in the more expensive instruments. The diaphragm is sometimes embossed with a *pattern of shallow troughs*, which by varying the depth, width, and number can be made to possess any degree of elasticity. Thus areas of stress are also reduced.

The fixed plate is mounted rigidly behind the diaphragm, and, in order to achieve as high a capacitance as possible very close to it. *Capacitance can vary from 5–75 pF, but between 20–30 pF is most common.*

There must be adequate passage of air behind the diaphragm, if the microphone is to work as a *pressure-gradient transducer*, so the whole area of the diaphragm cannot be mated with a corresponding area of backplate, while with others the plate is of lesser diameter than the diaphragm, thus giving an air gap around its edge. These passages open into the air chamber behind the fixed plate and from there through an acoustic-resistance element to the side vents.

For the capacitor microphone to function a voltage must be applied across it, and this also has the effect of pulling the diaphragm back towards the fixed plate by means of electrostatic attraction. Thus it is kept *taut* and *rigid*, a necessary characteristic if it is to respond faithfully to the incoming pressure waves, yet without any penalty in the form of increased mass.

The total capacitance of the unit is small, and therefore, *the capacitance variations are minute.* In order to produce current flows of usable proportions, the applied voltage must be high. In some cases it can be over 100 V, but around 50 V is common. Some models are designed to work well down to 9 V, and thus receive power from a standard radio battery. These will work with even lower voltages but the sensitivity tends to fall off.

Different voltages will give rise to different diaphragm tensions and sensitivities. As the acoustic resistance in the rear access passage is fixed in value, it follows that any such diaphragm variation will produce a difference in the way it responds to a *pressure gradient* between front and back. Thus the polar diagram will be modified. *Some models make use of this principle to achieve a wide choice of polar diagrams with the same microphone,* *ranging from omnidirectional, through figure-of-eight and cardioid to hypercardioid merely by altering the polarising voltage.*

To conclude, the *output level* of the capacitor microphone is extremely low, and a high-gain amplifier must be used with it. The amplifier should be mounted directly at the microphone, usually right in the microphone case. The capacitor microphone has very excellent frequency response and low distortion. *Because of the necessity of mounting an amplifier at or in the microphone case, the capacitor microphone is not recommended for ordinary high-fidelity (hi-fi) work.*

LECTRET MICROPHONES

In electret microphones, Fig. 2.15, *a permanent electrostatic charge is implanted into the metallised diaphragm, thereby eliminating the need of an external high-voltage source.*

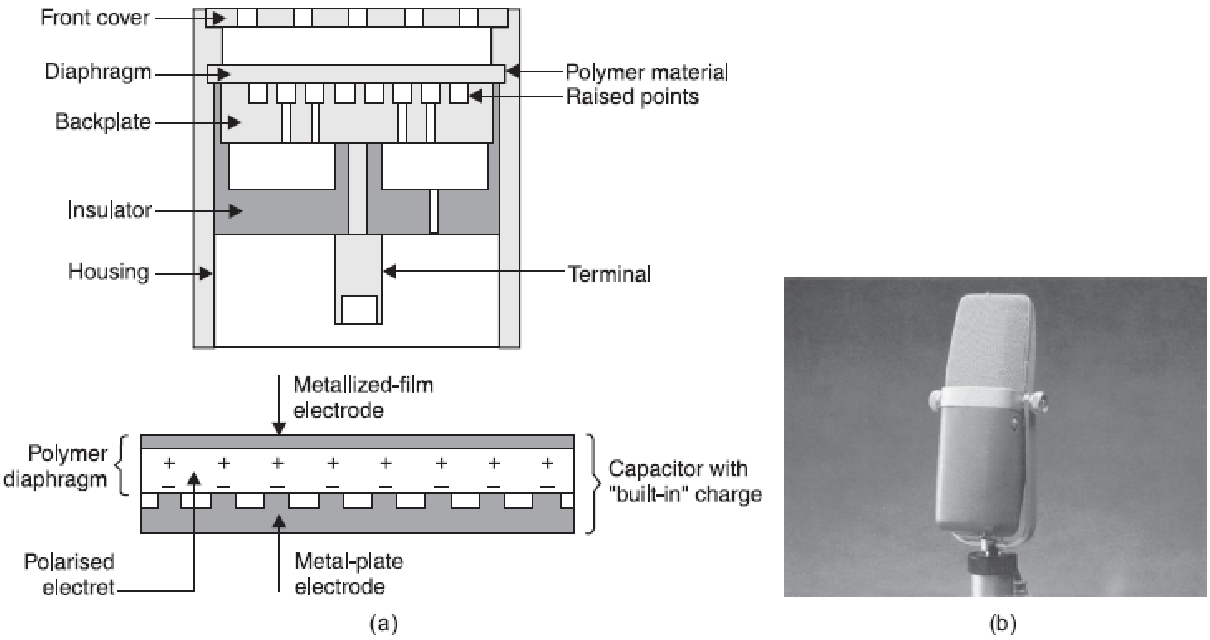


Fig. 2.15 (a) Construction details of electret microphone (b) Sony 6.38 B studio microphone

One way of doing this is to put the diaphragm sheet between the plates of an air-spaced capacitor which is then charged upto a high voltage. The sheet is then heated and allowed to cool with the charge still maintained across it. When it is removed, it has *a permanent static charge* which is the equivalent of an applied voltage of around 100 V. *This may be considered on par with a permanent magnet which retains its magnetism after the energising field has been removed.*

Any electrostatic charge will *slowly leak away* due to the fact that there is no material yet known that is a perfect insulator. The charge in the diaphragm will gradually diminish and also the effectiveness of the microphone. Its life is measured like that of radioactive substances, as time taken to fall to half its previous level, the *half-life* as it is usually termed. *The makers of* *electret microphones claim an expected half-life of between 100 to 1000 years, so users need have no worries on that score!* Conditions of high *humidity* though, may well accelerate the deterioration.

A *self-contained amplifier* is still necessary because the output impedance is of a similar order to that of the conventional capacitor microphones. This can be powered from a small battery cell. As the current taken by the amplifier is very low, just a fraction of a milliamp, it will have an *average life* of several thousand hours. This is in contrast to the battery life of capacitor microphones using dc converters to supply the polarising voltage which is in the lower hundreds.

Since their introduction, electret units have become very popular as they give reproduction approaching that of the conventional capacitor microphone but without the expense and complication of the external power unit. *Most of the built-in microphones fitted to cassette recorders are of this type.*

GUN MICROPHONES

The *necessity* of having a microphone within a short distance of the sound source can in some cases be an inconvenience. In particular, *with film and television work*, it is essential that the microphone is out of camera range, but there are many other applications where it is not possible to get very near to the source of sound.

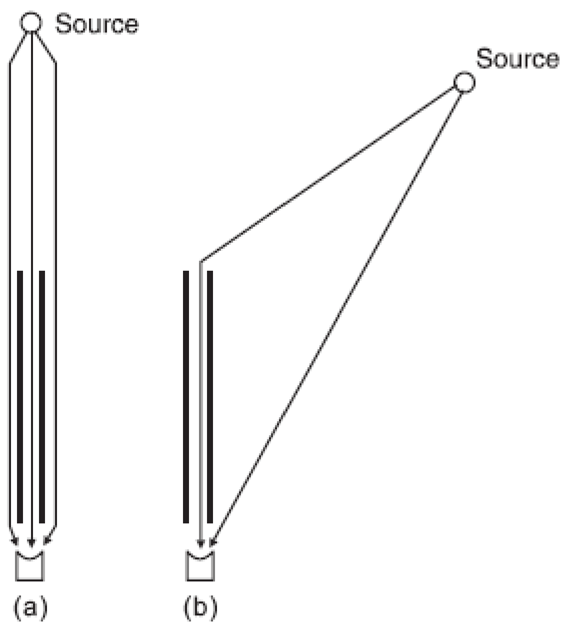


Fig. 2.16 (a) A single tube placed in front of a transducer, with a gap between them, directed toward a sound source. Sound travels down the tube to reach the diaphragm at the same time as that passing through the gap from outside. (b) When sound source is off axis, sound passing through the tube takes a longer path and so is delayed. When difference in path length equals a half wavelength of the sound there is cancellation.

An ordinary microphone will pick up sounds from quite a long distance, although the electrical output will be small because of the *square law reduction of sound intensity.* This in itself is no great drawback as extra gain can be provided in the amplifier. The real problem is that *unwanted sounds*, which normally would be so weak as to pass unnoticed, now assume a significant proportion compared to the *required sound.* The extra amplifier gain used increases them too. When operating *outdoors* this means that distant traffic, aircraft, dogs barking, wind noise and other such sounds become obtrusive, while *indoors* reverberation from walls and ceiling becomes greater in proportion to the direct sound, so giving an indistinct, distant effect.

For some while, professional engineers have been using the *gun* microphone, which is also known as the *rifle* and *barrel* microphone. *Sounds coming from all directions other than the one to which it is pointing are eliminated.* Two types of construction can be used, and both make use of the *interference* principle.

Let us first of all imagine a single tube mounted in front of a pressure microphone but with a *gap* between the diaphragm and the end of the tube. Sound coming from the direction of the tube axis will pass down the tube and reach the diaphragm *simultaneously* with the pressure wave that arrives from outside the tube through the gap [(Fig. 2.16(a)]. These will *reinforce* each other and so will actuate the diaphragm.

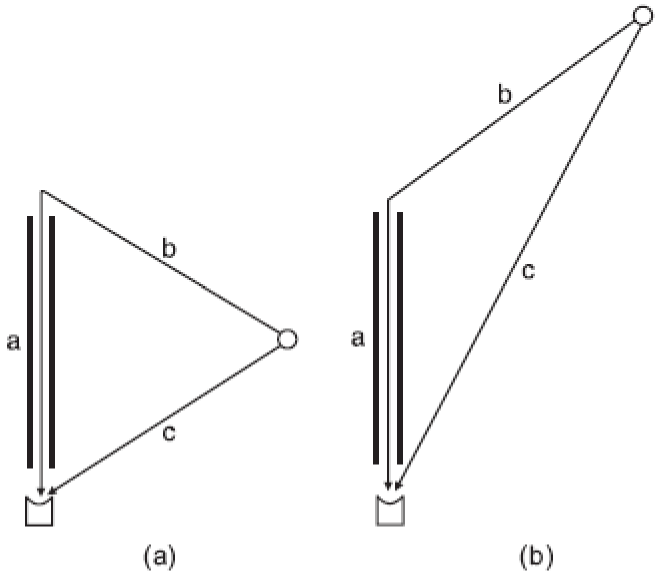


Fig. 2.17 Path-length difference varies according to angle of sound source At (a) b = c, therefore a + b – c = a; while at (b) b < c, therefore a + b – c < c. Thus the cancelled frequency rises as the source moves to a narrower angle.

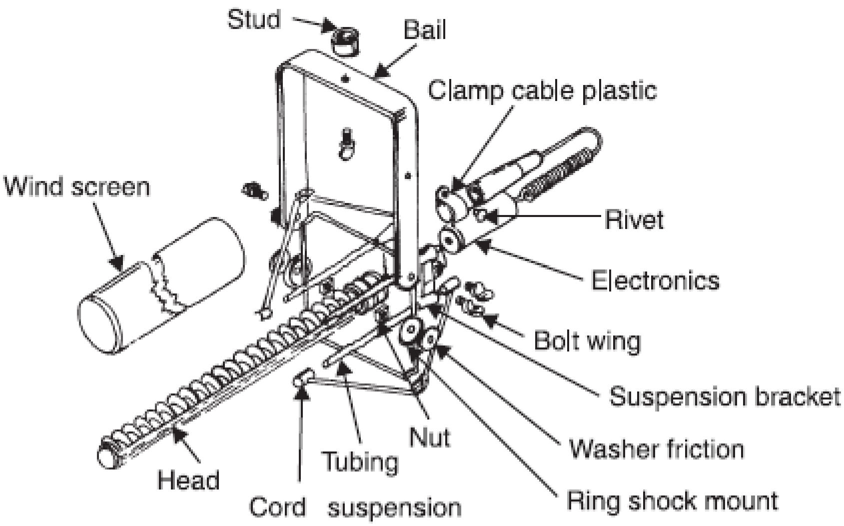


Fig. 2.18 The shotgun condenser microphone. Electro-Voice CL42S assembly exploded view

Consider now a sound coming from a point well off axis [(Fig. 2.16 (b)]. It follows two paths, one *direct* through the gap and the other *indirect* through the tube. The latter is longer and so the sound wave following this path arrives later. The difference in length between the two paths is equal to the length of the tube in the case of a sound source at right angles to the tube axis, but is less for any angle smaller than a right angle (see Fig. 2.17).

When this difference is equal to half the wavelength of the sound, there will be *half-cycle delay* between the waves following the two paths; *the compression part of one will coincide with the rarefaction part of the other.* Thus they will be exactly out of phase and will cancel each other. The result is therefore zero pressure on the diaphragm.

This cancellation occurs when the length difference is half a wavelength and so also at $1\frac{1}{2}$, $2\frac{1}{2}$, $3\frac{1}{2}$ wavelengths and so on. However, this can be true only when each successive sound wave is *identical* to its predecessors, such as in the case of a pure tone. Natural sounds are rarely like this as there are constant changes of wave shape and intensity, and so cancellation beyond more than a cycle or two will not occur. In practice, the half and perhaps $1\frac{1}{2}$ wavelength can only be expected to suffer cancellation. *As the path difference varies according to the angle of the sound source from the tube axis, the exact frequencies also vary.*

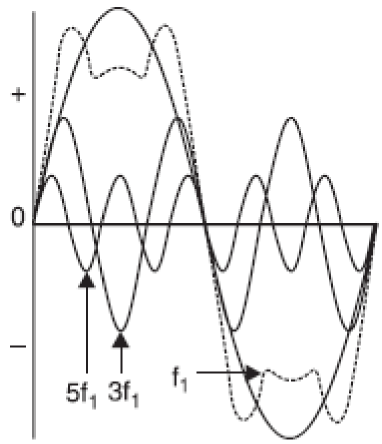


Fig. 2.19 Interview microphone

The advantage of gun microphone is to cut down sounds coming from directions other than the desired one, which means that the wanted sound is greater in proportion to the unwanted one, but it is not actually greater enough to be picked up by any other microphone.

Parabolic Reflector

A device which magnifies sound levels acoustically is the parabolic reflector. It works in a very similar manner to the way a concave mirror reflects light, but with additional complications. If a large concave surface is placed in the path of a sound pressure wave, the wave will be reflected and the direction of reflection will be governed by the angle of incidence between the wave and the surface.

In the case of a parabolic surface, the reflection from the whole surface area will converge on a single point which, as in the case of light, is termed the focal point. Thus the pressure is concentrated to a much higher intensity here. A microphone is mounted facing inward to the reflector, with its cone or diaphragm at the focal point, so that most of the sound reaching it is reflected from the parabola.

LAVALIER MICROPHONES

For many applications *mobility* is a primary requirement for a microphone. Gun microphones have some advantage over those permanently mounted on a stand as they can be used in the hand and carried around to the limit of the connecting cable length. As these can also be fitted to a stand with quick release clips, they have a versatility to which no doubt is due the popularity of this type of instrument.

A disadvantage is that *one hand of the user is always occupied and held up in an unnatural position.* This can impose a serious restriction on speakers or lecturers who may wish to demonstrate points with models or exhibits, use a blackboard or just be free to use gestures.

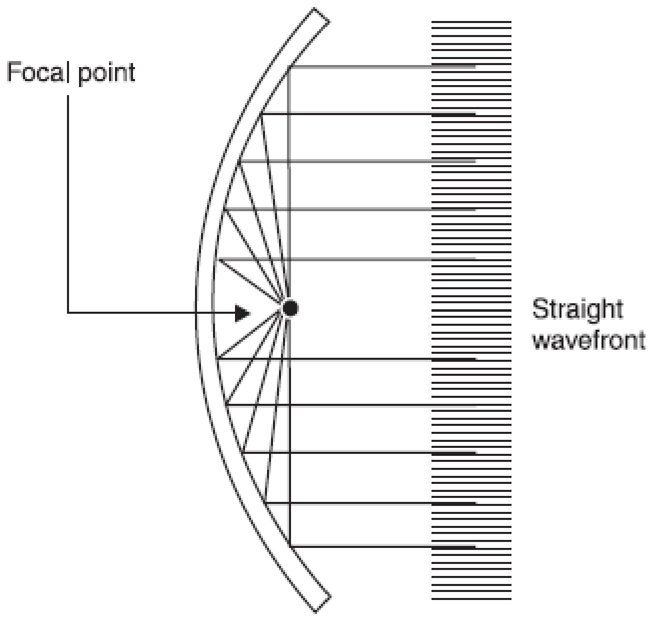


Fig. 2.20 Paths from a straight wavefront to a point of focus in front of a reflector must be of equal length so that sound arrives at the same time. This is ensured by using a parabola.

One answer to these problems is the lavalier microphone, which is *a small transducer suspended on the chest by means of a cord around the neck.*

Any type of instrument can be used for the purpose by arranging suitable fittings to take the cord. There are problems involved in using a microphone in this way. One of these is the *resonance of the human chest cavity* which emphasizes frequencies around 700 Hz and results in an unnatural boominess. Another is the *masking effect of clothing* which tends to absorb frequencies between 3 kHz and 10 kHz and give a muffled reproduction.

In order to overcome these effects special microphones have been developed for lavalier use. These have a *dip* in their response around 700 Hz and a *rise* from 3 kHz upward. These characteristics balance the deficiencies and result in a more or less flat overall response. The frequency curve in free air and that when used as a lavalier is shown in Fig. 2.21(a).

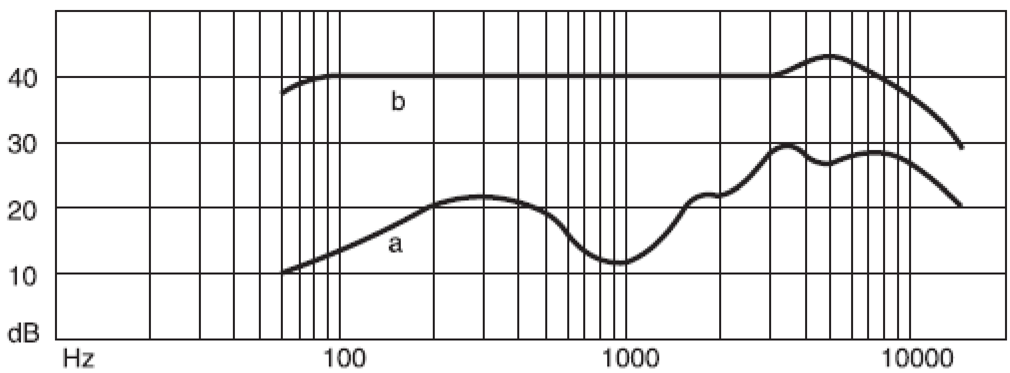


Fig. 2.21(a) Curve (a) shows the response of a typical lavalier microphone in free air but curve (b) gives the response when used as a lavalier. The trough at 700 Hz has equalised the chest resistance and the rising treble, the masking effect of the clothing.

In this respect an electret microphone would be ideal for this application as the frequency characteristic of most of these instruments is that of a rising treble and falling bass response.

Fig. 2.21 (b) Omni-directional microphone

TIE-CLIP MICROPHONES

There is another type of microphone with a similar purpose, but even more convenient, now gaining popularity. This is the *tie-clip or lapel unit* shown in Fig 2.22. This is a tiny microphone supported by a small mount which can be clipped on to a tie, lapel, or other convenient part of clothing. One of the prime requirements for such an instrument is *small size and lightness.* This rather cuts out the moving-coil transducer as there is a practical limit to the minimum size of cone/coil assemblies, also the need for a powerful magnet makes it too heavy for this application.

Here again *electret system* comes to its own. These units can be made very small and are inherently light. Although they need an *internal amplifier*, these can be formed on a tiny chip of silicon in the form of an *integrated circuit*, and so add hardly anything to the bulk or weight.

Fig. 2.22 Tie-clip microphone

A low voltage is needed to power the amplifier, but the current drain is very small, between 160-400 μA, so a small single cell can be used and still have an appreciable working life. Some microphones use a pen cell which is fitted in the microphone jack-plug, the current passing along the microphone cable, while others use a single mercury cell contained in the microphone itself. These are small disc-shaped cells, looking something like a *button*, and being only a few millimeters thick take up very little room. Battery life ranges from 5,000 to 10,000 hours or even more. In the case of capacitor transducers higher voltage and current is required, so this cannot be supplied from an internal battery but must be fed *along the cable* in some way.

As with the lavalier method of mounting, the frequency characteristic of the *electret* makes it very suitable for the purpose. *Chest cavity resonance* is less of a problem with the tie-clip or lapel unit because the microphone is not supported directly on the chest but stands off to some extent.

WIRELESS MICROPHONES

The ultimate in mobility is afforded by the wireless *(radio microphone)* because with this *there is no connecting cable and the user is free to move around over a distance of several hundred metres.* There are two basic types, one where the radio transmitter is contained within the casing of the actual microphone, and the other which takes the form of a slim pocket unit about the size of a wallet into which an ordinary microphone can be plugged.

The *integral microphone/transmitter unit*, Fig. 2.23, is rather larger than a normal gun microphone as batteries must be accommodated as well as the transducer and transmitter. In order to obtain sufficient power for the transmitter, the batteries are at least 9V, but the size limits the capacity. The average life is three to five hours, but *rechargeable batteries* are often fitted to make the instrument more economic to run.

With the *separate pocket transmitter* a lavalier or tie-clip microphone can be used to give complete freedom to the user. The aerial takes the form of a short flexible lead which trails from the microphone. Usual length is a *quarter wavelength* at the permitted frequencies of the carrier wave.

The transmission is picked upon a *special receiver* which tunes to the frequency used and demodulates the signal delivering an audio frequency output which can then be applied to an amplifier or recorder in the normal way.

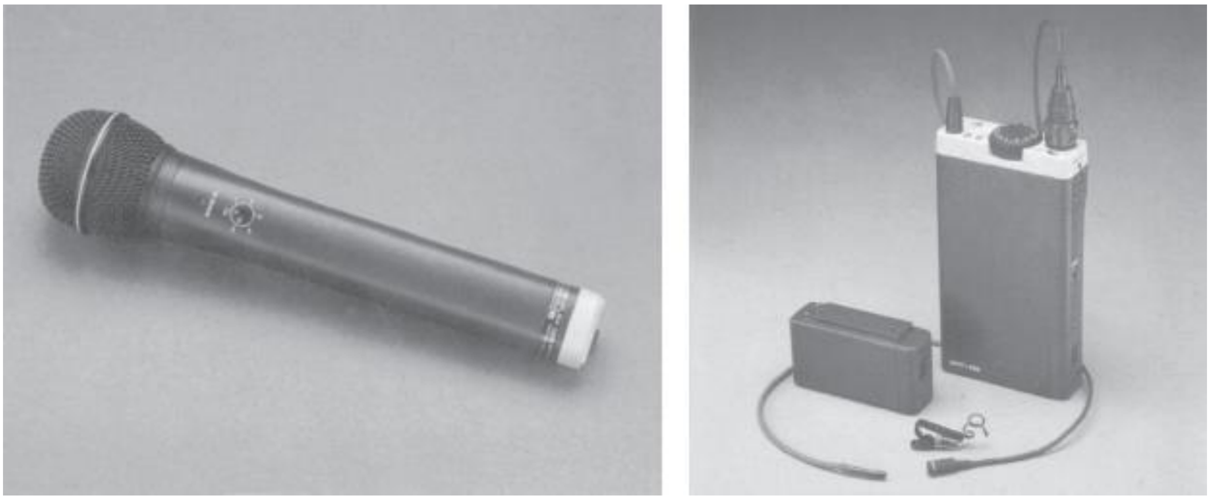


Fig. 2.23 (a) VHF wireless microphone (b) VHF transmitter

There are *fifteen frequencies* allocated for wireless microphones and all units work on any one of these. *Interference* is no problem because of the short range, it being unlikely that another user will be operating on the same frequency within about half a kilometer.

The frequencies are in *four groups:* firstly a group with a wide bandwidth, 174.1, 174.5, 174.8, and 175.0 MHz. The second group is of narrow bandwidth, the frequencies being 174.6, 174.675, 174.77, 174.885, and 175.020 MHz. The third group is also of narrow bandwidth, being reserved for teaching deaf children in schools; these are 173.4, 173.465, 173.545 and 173.64 MHz. In addition, in certain circumstances, the frequencies of 174.65 and 174.95 MHz are allocated for communication on work sites. An ordinary FM receiver will not pick up wireless microphone transmissions.

The *narrow bandwidth* specification is for a deviation of + 20 kHz, and is suitable for most speech applications. The *wide bandwidth* allocations allow a deviation of + 75 kHz and give the better quality reproduction required by stage and cabaret artists.

The transmitter *output power* must not exceed 50 mW in the case of narrow band transmitters, and 10 mW with the wide band units.

Certain specifications also apply to the receiver. *Signal to noise ratio* must be better than 30 dB and *selectivity* such that a signal with a deviation of + 10 kHz, 70 kHz away from the wanted signal in the case of *narrow band receiver*, and with a deviation of + 2.5 kHz, at 200 kHz away from the wanted signal in the case of *wide band receiver* will not produce an increase of noise plus unwanted signal of more than 3 dB in the output.

An *interfering signal* of 3 mV should not give a signal in the output greater than 10 dB above noise level in the case of wideband receiver and 20 dB above with that of the narrow band unit.

It is possible for any receiver to generate and radiate a signal from the local oscillator which is part of the superheterodyne circuit universally used. The specification stipulates that *any such signal radiated from the receiver's aerial should not exceed 2.5 μW at any frequency.*

These stringent requirements serve to protect the users from the effects of jamming and interference which undoubtedly would be common if such controls were absent. VHF wireless system is shown in Fig. 2.24.

Fig. 2.24 VHF wireless system

DUAL-UNIT MICROPHONES

Microphones are available which make use of two units to secure a *particular directional pattern*. A dynamic unit is often combined with a ribbon unit. Fig. 2.25 shows the directional pattern which results when a *bidirectional velocity unit* and a *nondirectional dynamic unit* are combined.

The resultant directivity pattern is a cardioid. Other units are also combined to secure specific directivity patterns.

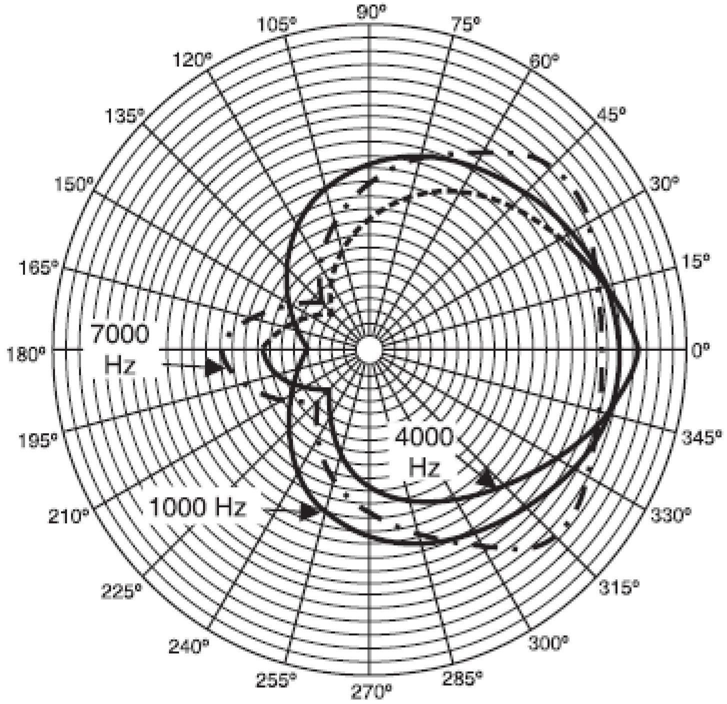


Fig. 2.25 Directivity pattern of a dual-unit microphone

A superb combination from Sennheiser is shown in Fig. 2.26. The MKH 30 is a pressure gradient microphone with a *figure-of-eight directivity*, optimising wide frequency response, lateral sound rejection and extremely low-inherent noise.

Fig. 2.26 A superb combination from Sennheiser

The MKH 30 is matched with the remarkable directivity of the MKH 60, a *supercardioid* microphone, to enhance the stereo-image. Low frequency ambience and vibration pickup is minimised by efficient rolloff filters.

Each microphone can be used *separately* when stereo operation is not required. Suggested applications of different types of microphones are given in Table 2.2.

Table 2.2 Microphones and their Applications

Type	Cost	Polar diagram	Operating principle	Quality	Suggested applications	Comments
crystal	low	omnidirectional	pressure operated; piezoelectric	low	Home recording system, Amateur communication, Mobile commu-nication	robust; outdoor use; high impedance
ribbon	medium/ high	figure-of-eight	pressure gradient; ribbon diaphragm in magnetic field	excellent; price a good guide	Drama, Music, Broadcast. In cardioid form it is suitable for gruop orchestra.	fragile; not for outdoor use; low impedance
moving coil	medium/ high	omnidirectional or cardioid	pressure operated; coil in magnetic field	can be very good	PA System, Broadcast, Music. In cardioid form, it is suitable for group orchestra.	robust; outdoor use; low or high impedance
capacitor	very high	omnidirectional, cardioid or figure-of-eight; can be switchable	pressure operated; diaphragm varies capacitance; polarising voltage.	excellent	Professional re-cording, Calibra-tion, Sound-level meters. Electret type is used in small halls, and clubs	fairly robust; outdoor use; norm. low impedance.

ELECTRICAL, MECHANICAL AND ACOUSTICAL ANALOGS

Any acoustical system can be represented in terms of an equivalent electrical or mechanical system. This is shown in Fig. 2.27.

The physicist freely uses these equivalents in setting up his mathematical approach in analysing a given system. For example, the effect of a cabinet on the functioning of a loudspeaker is clarified by thinking of the air in the enclosed space as acting like a capacitor in an electrical circuit, absorbing and giving up the energy imparted by the cone movements.

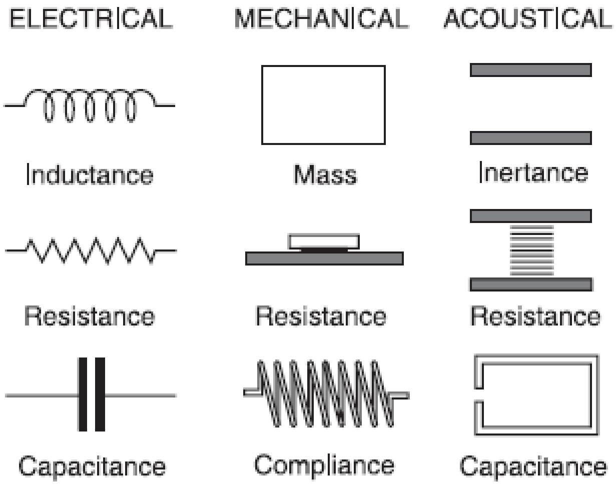


Fig. 2.27 The three basic elements of electrical system and their analogs in mechanical and acoustical systems.

Figure 2.27 shows *three basic elements* in electrical, mechanical and acoustical systems. *Inductance* in an electrical circuit is equivalent to *mass* in a mechanical system and *inertance* in an acoustical system. *Capacitance* in an electrical system is analogous to *compliance* in a mechanical system and *capacitance* in an acoustical system. *Resistance is resistance in all three systems*, whether it be frictional loss offered to air particle movement in glass fibre, frictional losses in a wheel bearing, or resistance to flow of current in an electrical circuit.

Descriptive Questions

1. What are the characteristics of microphones? Explain.
2. Explain the significance of a polar diagram.
3. With the help of a diagram explain the working of a dynamic microphone.
4. How does a ribbon microphone differ from a dynamic microphone?
5. How does an electret microphone differ from a capacitor microphone?
6. What are the limitations of crystal microphones?
7. Explain the significance of a hum-bucking coil.
8. In which microphones is a built-in transformer a necessity and why?

Multiple Choice Questions

1. Microphones are electroacoustic transducers which convert
    a. acoustic energy to electrical energy
    b. electrical energy to acoustic energy
    c. acoustic energy to mechanical energy
    d. mechanical energy to acoustic energy
2. The unit of sound pressure used for rating microphones is
    a. pascal
    b. bar

c. watt
d. decibel

3. The polar diagram of a true omnidirectional microphone is
   a. perfect circle
   b. figure-of-eight
   c. cardioid
   d. none of the above
4. Moving coil microphones are
   a. active transducers
   b. unidirectional
   c. both (a) and (b)
   d. neither (a) nor (b)
5. Moving coil microphones are
   a. omnidirectional
   b. unidirectional
   c. (c) neither (a) nor (b)
6. Carbon microphones
   a. produce an emf
   b. modulate the current from an external battery
   c. both (a) and (b)
   d. neither (a) nor (b)
7. The natural impedance of a ribbon microphone is
   a. very low
   b. low
   c. high
   d. very high
8. The ordinary dynamic microphone is
   a. omnidirectional
   b. unidirectional
   c. nondirectional
   d. either (a) or (b)
9. The polar diagram of a capacitor microphone can be modified by
   a. varying the polarising voltage
   b. varying the pressure gradient
   c. both (a) and (b)
   d. neither (a) nor (b)
10. The half-life of electret microphones varies from
    a. 100 to 250 years
    b. 250 to 500 years
    c. 100 to 1000 years
    d. 250 to 1000 years
11. The construction of gun microphones makes use of the principle of
    a. diffraction
    b. interference
    c. reflection
    d. refraction
12. A device which magnifies sound levels acoustically is the

a. lavalier microphone
b. gun microphone
c. parabolic reflector
d. ribbon microphone

Fill in the Blanks

1. For most high quality microphones impedance is ___________.
2. Cassette recorders of good quality are not, as a rule, designed to take ___________ impedance microphones.
3. Microphones do not respond equally to ___________ coming from all ___________.
4. An ___________ microphone is most commonly used for general recording.
5. An inset pattern of carbon microphone is in general use for ___________.
6. The carbon microphone generates a continuous ___________.
7. The double crystal unit is called a ___________.
8. The equivalent circuit of a crystal microphone consists of a ___________ in series with a ___________.
9. One of the refinements in better quality dynamic microphones is ___________.
10. The output voltage of a ribbon microphone is ___________ and so also is its ___________.
11. In order to produce large capacitance variations the ___________ must be high.
12. A self-contained amplifier is necessary with ___________ microphones.

ANSWERS

Multiple Choice Questions

1.	(a)
2.	(b)
3.	(a)
4.	(a)
5.	(a)
6.	(b)
7.	(a)
8.	(c)
9.	(a)
10.	(c)
11.	(b)
12.	(c)

Fill in the Blanks

1.	low
2.	high
3.	sounds, directions
4.	omnidirectional
5.	telephone systems
6.	hiss

7. bimorph
8. voltage source, capacitor
9. a hum bucking coil
10. very low, impedance
11. polarising voltage
12. ribbon

CHAPTER 3

HEADPHONES AND HEARING AIDS

Headphones offer the possibility of the ultimate private listening experience. Their small transducers can reproduce extremely low-frequency tones when connected almost directly to the ear, and their small size contributes to good high-frequency response. The privacy (therefore, low acoustic leakage), extreme bandwidth, and dynamic range possible at the ear make the headphone a vital professional tool and a relatively low cost, super-fidelity personal listening system.

HEADPHONES AND HEADSETS

A *headphone*, Fig. 3.1, is defined as a listening device consisting of either one or two earphone receivers and a headband to hold them in place. A *headset*, Fig. 3.2, is essentially the same thing with a microphone attached to it. Headphones permit *one-way* communication while headsets permit *two-way* communication.

Particularly since the introduction of *stereo*, the use of headphones has become an important part of the high fidelity *(hi-fi)* scene, in some cases *by choice* and in others by *necessity*.

In the early days of wireless, headphones—often described as *earphones*—provided the most practical way of listening to the then available signals. From 1923 onwards, with the emergence of public broadcasting and the availability of more powerful receivers, loudspeakers rapidly gained favour especially for family listening. Headphones, however, were still widely used with professional and communication equipment until well after World War II.

Fig. 3.1 Headphones permit one-way communication

Fig. 3.2 Headsets permit two-way communication

TYPES OF HEADPHONES

Headphones are available in a vast assortment of shapes, sizes and types. Advertisements emphasize distinctive features even further, to the point where it is understandable that consumers get a little confused about the importance of *features and forms.*

In headphones, as in almost anything else, most of the differences exist merely as evidence of the manufacturer's struggle to be distinctive, to market unique products.

One can identify *three basic classes* of headphones: (1) supra-aural (2) circum-aural, and (3) intraaural.

(1) Supra-aural (Fig. 3.3): Those that sit on the ears, usually pressing the external ear to the side of the head. Old versions of supra-aural headphones presented an uncomfortably hard bakelite or hard rubber surface to the ear. Nowadays *soft cushions are the rule*, some are foam or liquid filled and attempt an *airtight seal* to the irregularly shaped external ear. Others are fitted with large pads of acoustically transparent plastic foam and are designed in such a way that an airtight seal is not necessary.

Fig. 3.3 A compact pair of headphones with adjustable headband and soft foam pads for comfortable listening

(2) Circum-aural (Fig. 3.4): Those with cushions that fit around the external ear, pressing directly against the head itself. In the best examples of circum-aural headphones the external ear is not deformed through contact with the unit.

(3) *Intra-aural* (Fig. 3.5): These are in-the-ear or *insert earphones* of the type found in some aircraft, almost all hearing-aids and that are supplied for private listening to portable radios, tape recorders and TV's.

These classes describe headphones by their construction and how they are worn, but they tell us little about factors important to stereo listening.

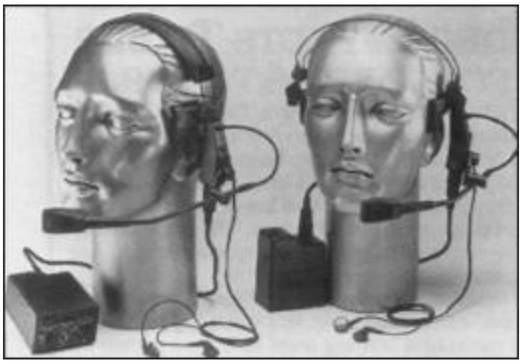


Fig. 3.4 Circum-aural type fit around the external ears

Fig. 3.5 Intra-aural (in-the-ear)earphones

One important distinction among headphones can be made on the basis of *whether they exclude external noise* and *prevent leakage of music to the outside world.* The former is important for satisfactory listening in a noisy environment; the latter might be important in hospitals, libraries, or perhaps at home, where even the low level leakage of music can be a distraction to someone else.

In order for a headphone to have *effective acoustic isolation* it must provide:

1. an *airtight seal* with the external ear (supra-aural) or the head (circum-aural) and
2. a *totally enclosed cup* over the ear, made of a rigid and relatively dense material.

These requirements eliminate headphones with electrostatic or electromagnetic membrane drivers (open backs) and porous cushion types (no airtight seal). *Insert (intra-aural) earphones that fit the ear canal singly can be very effective at acoustic isolation.*

MOVING-IRON HEADPHONES

The so-called magnetic type (moving iron) headphones use the same basic principle as was originally adopted for telephone earpieces or receivers. As shown in Fig. 3.6 a thin flexible soft iron *diaphragm* is supported by the casing just clear of the tips of two soft iron *pole-pieces* which are attached, in turn, to a small *permanent magnet.* Magnetic attraction causes the diaphragm to be distended slightly towards the pole tips, while still leaving a clearance of something less than a half-millimeter.

The incoming audio current passes through coils wound around the pole-pieces. In the presence of an audio signal the *varying magnetic field* created by the current *interacts* with that from the *permanent magnet,* varying the *pull* on the diaphragm and causing it to *vibrate,* thereby *generating* corresponding sound waves.

Because of the inherent *mass and stiffness* of the metal diaphragm, headphones of this type suffer similar *limitations* to the old horn-type magnetic loudspeakers; a *natural resonance* at a few hundred hertz, resulting in the so-called '*metallic quality*'. They also exhibit quite *high harmonic distortion* due to the fact that the diaphragm drive is one-sided or unbalanced.

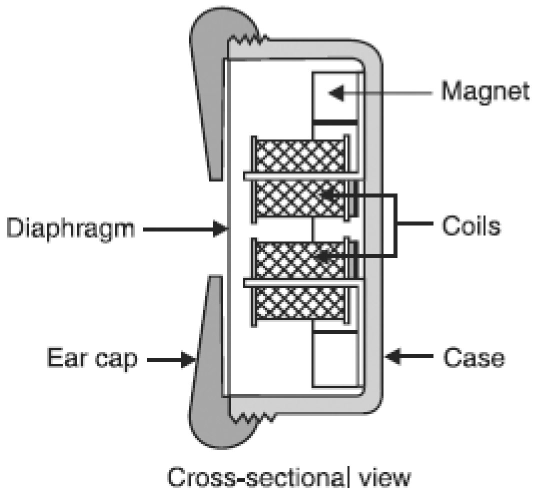


Fig. 3.6 Though refined over the years, moving-iron headphones still use the same principles as seen in the earliest telephone 'earpieces'. They have no place in today's hi-fi scene.

CRYSTAL HEADPHONES

In the 1950s and 1960s an assortment of crystal earphones appeared on the market, being essentially miniaturised counterparts of the crystal (or *piezoelectric*) loudspeakers.

Involving few metallic components, they could be made very small and light, as well as lending themselves to mass production. Many, infact, were designed to plug directly into the ear canal, being intended for use with portable radio sets and tape players. A few special quality models have been released over the years for monitoring situations but in general, *crystal headphones have slotted into a utility role rather than as a medium for hi-fi listening.*

As with crystal loudspeakers, *their impedance is predominantly capacitive*, measuring many thousands of ohms at low and mid frequencies. This may not present a problem where sufficient signal drive voltage is available, but *with ordinary low-impedance drive circuits, their effective sensitivity is likely to be poor compared with magnetic units.*

DYNAMIC HEADPHONES

The widespread adoption of dynamic or *moving-coil* loudspeakers in the 1930s raised the question as to whether it would be possible to use the same principle for headphones. If the *drive units* could be made small enough and light enough, they would presumably offer equivalent frequency response and much reduced distortion.

Moving-coil earphones designed in 1936 and released in 1937 were discontinued in 1939 as *too cumbersome and too expensive* to justify continued production. But by 1967 twenty seven firms had entered the market and one-time descriptions like *heavy, cumbersome, and metallic* had also given way to *lightweight, comfortable, and high quality*, with headphone listening rapidly becoming an accepted part of the hi-fi stereo scene.

Because it was the most accessible approach, many of the dynamic headphones introduced to the postwar hi-fi market comprised, essentially, of two miniature (50-70 mm) loudspeakers housed in *shells* rather like the two halves of an electromagnetic unit egg.

Typically, the oval shells of a pair of middle-aged *Japanese-made Golding dynamic headphones* measure about 10 cm tall, 9 cm wide, and 5 cm deep. They are fronted by a gold anodized *grille* with two matching *pressure relief ports* at the rear. The *shells* are cushioned with a removable pad, while the *headband* is adjusted and padded. While bulky and by no means light, the treatment gives them a soft and friendly feel.

One can reasonably expect that equivalent models from reputable sources would incorporate drivers with magnets, cones and voice coils selected for the role and *with the shells physically styled and acoustically treated for the best available end result.* With outside noise largely excluded, and operating so close to the ears, the drivers can be expected to produce an ample *sound pressure level (SPL)* with only milliwatts of input thereby *minimising cone break-up and non-linear distortion.*

Being *damped* by both the front and rear air loading, cone resonance effects are held in check, flattening frequency response. Subjectively, *high frequencies are projected directly into the ear canal while the captive air inside the cushion surround also conveys low frequency pressure waves efficiently to the eardrums.* Most such headphones are capable of sound quality far better than is available from traditional magnetic earphones.

While such headphones are still in use, many people find them too cumbersome for prolonged listening and, in a warm environment, too hot around the ears. *Now-a-days most stereo headphones are much smaller and lighter, using specially developed ultra-compact drivers. Some retain paper based cones; others retain a dome type diaphragm similar in configuration to the squawker and tweeter loudspeakers. A few up market models include their own diminutive tweeter.* Dynamic headphones are shown in Fig. 3.7.

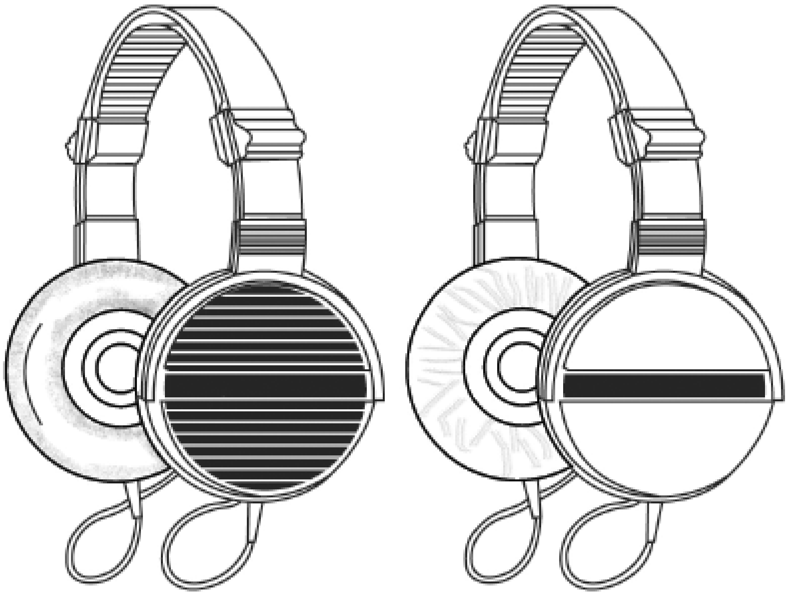


Fig. 3.7 Dynamic headphones with a very low mass of the diaphragm and moving-coil assembly

ELECTROSTATIC PHONES

Another type of headphone needs a special mention at this juncture—those using the capacitive or *electrostatic principle.*

A very thin, partially conductive plastic membrane, is supported just clear of a metallic surface being, at the sametime, attracted towards it by an electrostatic charge. *An audio signal voltage, applied to either element, adds to or subtracts from the charge, causing the membrane to vibrate and produce sound waves.*

This obviously parallels the earlier explanation of a magnetic earphone (Fig. 3.8) but with the vital difference that *a plastic membrane can have much less mass and stiffness than a metal diaphragm, thereby minimising problems with system resonance and uneven response over the working range.* It also lends itself to a *balanced configuration* minimising harmonic distortion.

In early model electrostatic headphones, as with their headphone counterparts, a *high dc polarising voltage* had to be fed to the phones along with a *fairly high audio drive voltage.* This meant, in practice, that electrostatic headphones could be simply plugged into a socket provided for dynamic headphones. An *external adapter* was required, containing a mains powered high voltage supply and a step-up transformer for the audio drive signal.

ELECTRET ELECTROSTATIC HEADPHONES

In the mid-1970s *permanently polarised* or electret elestrostatic headphones made their appearance, with *Sennheiser* Models 2000 and 2002. While the need for a mains powered h.t. supply had been eliminated, they still need an external black box containing *twin step-up transformers*, to permit operation from the amplifier voice coil output terminals.

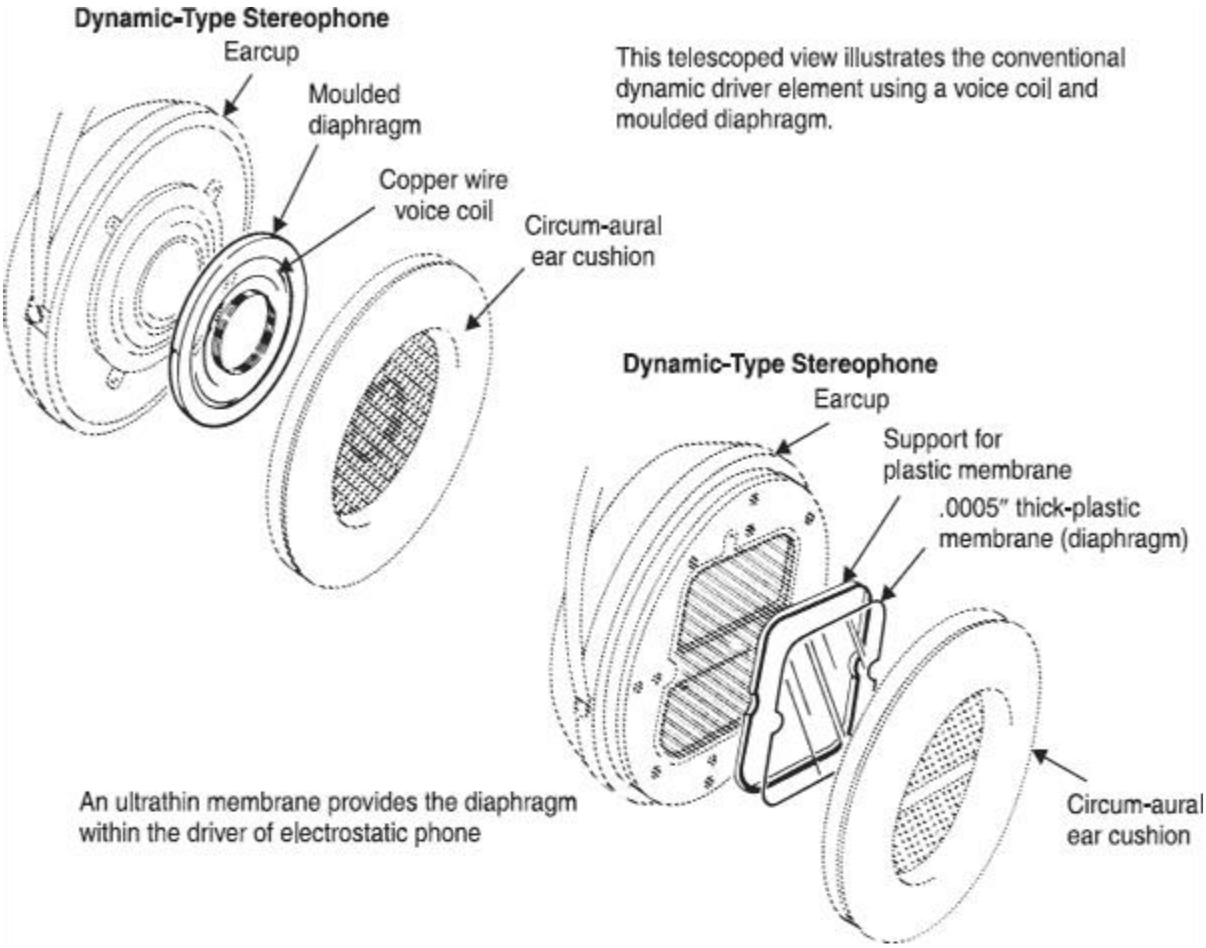


Fig. 3.8 Difference between dynamic and electrostatic headphones

In fact, Sennheiser made a virtue out of necessity by arranging for the adapter to feed two separate pairs of electrostatic headphones as well as providing *loudspeaker/headphone switching.* The company also included *protection circuitry* to limit the sound pressure level from the phone to 117 dB, alongwith *l.e.d. indicators* to show when the circuitry was being activated.

About the same time, AKG brought out its two-way 4K headphones which, despite their modest dimensions, were a *two-way design combining a dynamic driver and electret tweeter crossing over at 4.5 kHz.*

As indicated in Fig. 3.9, the *woofer* is deep inside the casing. Using a *highly efficient rare earth/cobalt magnet* and a *moulded polycarbonate diaphragm*, it can operate directly from a normal amplifier headphone socket. Its rated *low end response* extends to 16 Hz but, at the top end the body of the tweeter serves to block physically its residual output above the *nominal crossover frequency*.

The *tweeter* uses the electret principle, with a 12.5 micron chromium-sputtered foil optimised for the frequency range above 4.5 kHz. The higher drive voltage required is provided by a 1:5 step-up transformer, which can be made extremely small by virtue of the fact that it needs to handle only frequencies above 4.5 kHz, falling off steeply below that.

For the design, AKG claimed the best of both worlds—*the ability to operate from a normal headphone drive circuit and the reputed sonic transparency of an electroacoustic transducer for the top end.*

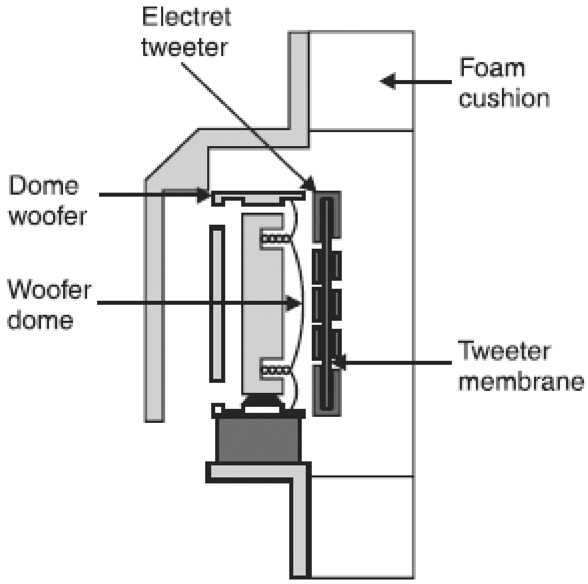


Fig. 3.9 Headphone combining a dome type dynamic woofer with an electret electrostatic tweeter, fed by a miniature step-up transformer. It can operate directly from a normal dynamic headphone feed socket.

HEARING IMPAIRMENTS

There are millions of people in the world with hearing impairments. It is surprising that the special hearing needs of so large a group have been largely ignored for so long, especially when each of us faces the very real probability of joining that group through disease, trauma, or just by growing old.

The hearing impaired are not just the people who wear aids. In fact, only about 20% of the hearing impaired wear aids. *Many people with hearing losses are able to function in close or face-to-face situations but are lost in noisy or reverberant settings.* Even people who wear hearing aids have problems in reverberant rooms or where there is a high background noise level. Our standards for speech intelligibility are based upon *listening tests* with normal hearing subjects and are not directly applicable to the hearing impaired. Noise and reverberation degrade intelligibility far more rapidly for the hearing-impaired individuals—whether they are fitted with hearing aids or not. Often the very highly prized acoustical qualities of our theatres and concert halls operate against the needs of the hearing impaired. The acoustical design of classrooms and lecture halls are almost always inadequate for the needs of the hearing-impaired student.

In recent years whole new technologies have been developed or adapted to meet the special needs of the hearing impaired in public assembly spaces. Each of these new systems has one thing in common—they are *wireless.*

AUDIOMETRY

Audiometry is the general term applied to a number of tests used to measure hearing.

An audiometer is an essential instrument for conducting most audiometric tests. *The audiometer is basically a device that generates pure tones of known level and frequency, and delivers these to the patient through either a headphone or a bone conduction vibrator.*

On an audiometer, the levels are given as *Hearing Level*, dB *(HL).* In this way, a normal hearing person will be able to hear tones at and above a 0 dB *HL* level.

Since these levels are average levels measured in a standard coupler, the actual sound levels at the patient's eardrum can differ considerably. One might expect up to as much as 15 dB individual *deviation* from the average.

Instead of using headphones one can use *insert earphones*, where the transducer, by means of an earplug, is inserted into the ear canal. Insert earphones offer some advantages regarding reduction of possible *masking effects* from environmental noise during audiometry. Also in the case of *non-symmetrical hearing loss*, insert earphones are preferable, as their coupling to the opposite ear is minimal. Finally, insert earphones can be used in cases where the ear canal collapses when using conventional earphones.

When recording the *air conduction threshold* the sound must pass through the ear canal and the middle ear to reach the hair cells in the inner ear. When the *bone conduction threshold* is recorded, a bone conductor is used. The signal then by-passes the ear canal and middle ear and is conducted directly through the bones of the cranium to the inner ear. Therefore, if the air conduction threshold is worse than the bone conduction threshold (an “air-bone gap”), the reason must be found in the ear canal or in the middle ear.

An *audiogram* contains information about the hearing loss at a number of frequencies. *The total hearing loss, as measured with headphones at any frequency, is called the air conduction hearing threshold level. The sensory part of the hearing loss, as measured with a bone conductor, is called the bone conduction level, or bone conduction threshold.* Note that not all audiometers have a bone conductor facility.

The audiogram is a graphic representation of the audiometric results. Fig. 3.10 shows an example of an audiogram.

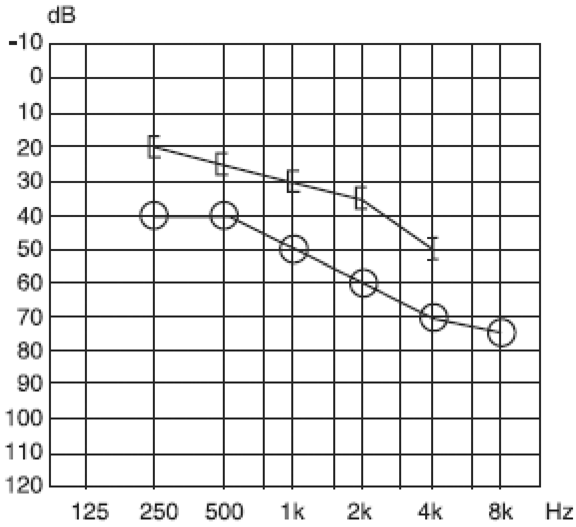


Fig. 3.10 Right ear audiogram measured with both air (O) and bone (I) conduction. The average air-bone gap is about 20 dB indicating a mixed hearing loss. The air conduction symbol for the left ear is (X) (not shown).

Air Conduction

The Hearing Threshold (also referred to as the Hearing Threshold Level, *HTL)* is an important audiometric measurement. *HTL represents the weakest sounds the person can hear during the test.*

Pure tones are presented to the test person through headphones and he or she must indicate when a tone is heard. By varying the level we can determine the *hearing threshold* at any test frequency. The values are plotted on an audiogram, which contains data for at least the frequencies 500 Hz, 1 kHz. 2 kHz and 4 kHz. A more complete audiogram covers also 125 Hz, 250 Hz, 1.5 kHz, 3 kHz, 6 kHz and 8 kHz. If the two ears have different thresholds, the *good ear* should be tested first. If the hearing loss on the *bad ear* is 40 dB or more below the (bone conduction) level of the good ear, masking should be used to cancel out the effect of bone conducted sound.

A source of measurement error is *background noise.* This will cause the threshold to be worse, due to masking of the test ear. Therefore, *audiometry is normally carried out in a sound treated test booth.*

Bone Conduction

It is important to distinguish between *conductive losses* (in the outer ear or in the middle ear), *sensorineural losses* (in the inner ear, in the nerves or in the auditory pathway) and *mixed losses* (a combination of a conductive and a sensorineural loss).

For this purpose, the bone conduction threshold is often included in the audiogram. For measuring the *bone conduction threshold*, a small vibrator is usually placed on the scull (mastoid) just behind the ear, instead of the headphones. This vibrator is called a bone conductor. The *bone conductor* should not touch the pinna during testing, as vibrations could be sensed from its casing, causing a false result.

The difference between the bone conduction and air conduction thresholds is called the **air-bone gap.** The extent of the air-bone gap is a measure of a possible conductive component of the hearing loss. In a sensorineural loss, no air-bone gap is present. Small differences, up to 10 dB, can be considered insignificant. Such differences can be the result of individual variations in the coupling between the vibrator and the scull. The average air-bone gap (a measure frequently used in the fitting process) can be calculated as the **average** of the air conduction and bone conduction differences over the frequency range evaluated. An air-bone gap of 15 dB or more, indicates a true conductive component.

Distortion in the bone conductor renders the results at low frequencies less reliable than the air conduction measurements. With a *flat* hearing threshold curve, only the high frequencies are valid. With *steeply sloping* curves, a larger air-bone gap is necessary to make sure a true conductive loss is present.

The sound from the bone conductor can often be heard equally well in both ears. The bone conduction threshold is therefore a measure of the best ear at that frequency. If the bone conduction threshold for one ear only is to be measured, the opposite ear must be masked with noise.

Un-Comfortable Level (UCL)

The *Un-Comfortable Level, UCL,* **refers to the level above which the patient will not accept listening.** UnComfortable Level is also referred to as Threshold of Discomfort *(TD)* or Uncomfortable Loudness Level (*ULL*).

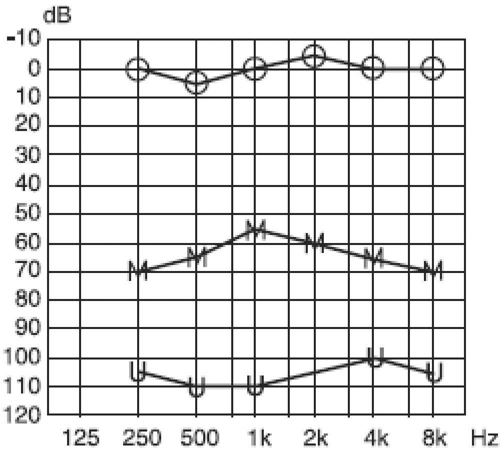


Fig. 3.11 (a) HTL, MCL and UCL for a typical normal hearing person

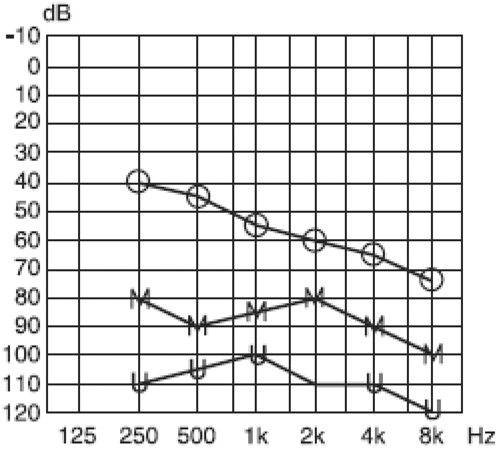


Fig. 3.11 (b) HTL, MCL and UCL for a typical hearing impaired person

This measurement is relevant to both diagnostic and hearing aid fitting purposes. Together with the hearing threshold, the UCL gives an indication of the total hearing range, the dynamic range.

Most Comfortable Level (MCL)

The level at which a person prefers to listen to a particular sound for a long period of time is called the Most Comfortable Level (MCL). MCL can be recorded for pure tones and for complex sounds such as speech and noise. When an MCL is recorded on several occasions, the variation observed may amount to 10 – 15 dB, or even more. Often, an MCL is estimated without measurement to be midways between the hearing threshold and the uncomfortable loudness level. MCLs are commonly used in hearing aid fitting to estimate the most suitable gain requirement.

Masking

When two ears differ substantially in hearing sensitivity, problems can arise when testing the poorer ear. The problem is that intense sounds, which are necessary to test the worse ear, may be detected by the better ear. Sound waves can give a sensation of hearing through two different pathways. The most efficient of these pathways is through the ear canal, the middle ear, and on to the cochlea. Also, sound vibrations can travel through the skull to the cochlea. If one ear is better than the other, it risks picking up vibrations caused by the loud sounds used to test the poorer ear. This may cause the patient to respond when the sound is still too weak to be heard in the worse ear. Masking is used to eliminate this problem.

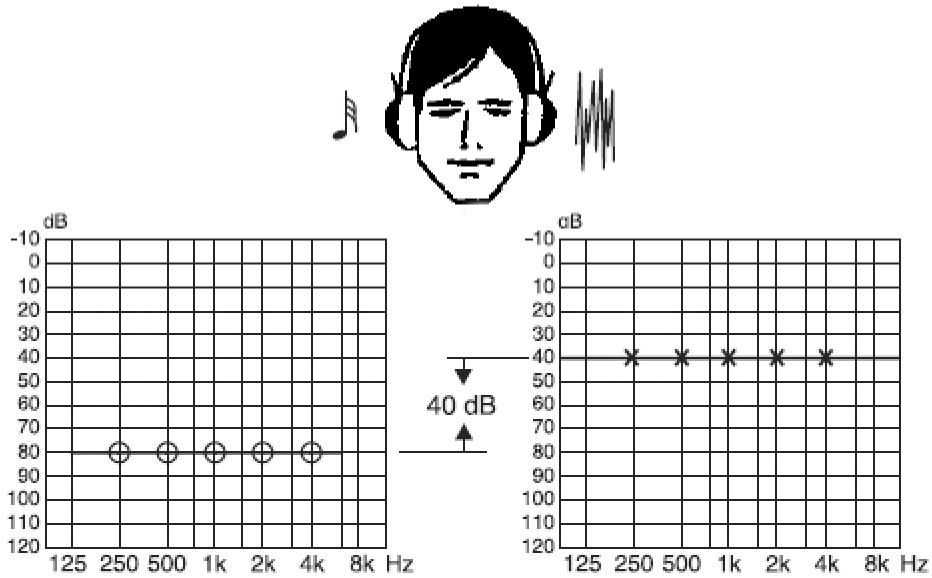


Fig. 3.12 The application of masking to the better ear when testing the poorer ear

Masking is the presentation of a noise signal to the better ear sufficiently loud to prevent the test stimulus from being heard in that ear. As a rule of thumb, masking is applied in air conduction measurements if the difference between the two ears is 40 dB or more. With a conductive hearing loss there are special problems. Masking should then be applied to the better ear when its bone conduction threshold differs more than 40 dB from the worse ear's air conduction threshold.

When the poorer ear is tested, a masking noise signal from the audiometer is presented to the better ear. By varying the level of the masking noise, it is possible to register whether the tone is heard in the test ear (the poorer ear), or whether the sound is erroneously detected by the opposite, better ear.

Speech Audiometry

In speech audiometry a number of test words or sentences are presented to the patient. The speech signal may be presented through earphones or, when e.g. a hearing aid is to be evaluated, through loudspeakers. Presentation through loudspeakers is called sound field testing.

The Speech Reception Threshold *(SRT*) is recorded by varying the level of speech until a certain proportion of the test words is correctly identified. *SRT* is typically the level at which the patient can understand 50% of the words presented. Normally, material consisting of numbers or two-syllable words is used for assessing *SRT.*

Very often, speech testing is carried out at a higher presentation level, often at the most comfortable level. The percentage of correctly identified test words is called the “speech recognition score.” Often, the older term “speech discrimination score” is used instead. Also the term “speech recognition loss” may be used, meaning the difference between a perfect score of 100% correct and the score actually obtained.

Speech audiometry can be used to compare the effect of different hearing aids or different settings of the same hearing aid. Caution should be taken in interpreting such scores, since they may differ substantially from test to test just by chance. A number of factors other than those related to hearing may influence the scores. Such as training, education and even dialectal background of the patient.

Impedance Audiometry

Impedance audiometry offers a convenient means of registering the status of the middle ear. A low frequency tone is presented in the closed ear canal. The level of the tone will be higher when the eardrum is **rigid** (immobile), compared to when it is **compliant.**

By this means it is possible to measure the static pressure in the middle ear, the integrity of the middle ear ossicles, whether the middle ear cavity is filled with fluid or air, and whether the middle ear muscles are activating or not (the *acoustic reflex*).

HEARING AIDS

The basic function of a hearing aid is to amplify sound in such a way that it is made audible to the hearing aid user. Although many types of hearing aids are available, they all share a number of common features. We will look in some detail at the various hearing aid types, their components and electroacoustic characteristics.

Full Concha Model

The full concha model will occupy most of the concha of the outer ear. It can provide more amplification and output than the canal model. A variation of this model is called a “semi” or “half' concha, in which the helix part, or most of it, is removed.

Canal Models

Canal aids, especially in their CIC format (Completely In the Canal) are the smallest type of hearing aid available. Because of its small size, only miniature batteries and receivers are used, causing limitations in amplification, output and battery life. The receiver and microphone are placed so close to one another that the risk of feedback is higher than in other models. Therefore, canal aids are used mainly for mild-to-moderate hearing losses.

Body Worn Aids

In most instances, body worn aids are used for those having severe to profound hearing losses. In some cases, they are used when the user has difficulty in handling other types of hearing aids. Such may be the case for the very young or very old hearing impaired person.

A body worn aid can provide more amplification than other hearing aid types because the receiver is separated from the hearing aid by a cord. This reduces the risk of **feedback**, permitting more power to be utilised. The receiver is snapped onto the earmould. The body worn aid is worn in a pocket or can be hidden by clothing. Very young children often use a microphone cover to prevent food from damaging the microphone.

Hearing Glasses

In hearing glasses the amplifier, microphone and receiver are built into the side arms. This was for many years a popular way of camouflaging a hearing aid and some users found hearing glasses easier to manipulate than BTEs. However, the practical drawbacks of combining visual and hearing rehabilitation into one device have become so evident that hearing glasses are now slowly disappearing from the market.

A variation of hearing glasses is the use of adapters which connect a BTE hearing aid to the side arms of spectacles.

CROS/BICROS

Hearing glasses are often used in a CROS application (Contralateral Routing Of Signals), where the spectacle frame is used to conceal the wires. CROS applications are sometimes used to fit *asymmetrical losses.* The microphone (in an empty hearing aid case) is mounted on the side having the more severe hearing loss. The amplifier and the receiver are mounted on the other side. The signal from the microphone is led to the amplifier, either via a cord or using wireless transmission. In most cases, it is possible to use an open mould.

A simple CROS application is used when the user has a severe hearing loss, or no hearing at all, in one ear and close to normal hearing in the other. A slightly more complicated application is the *BiCROS* fitting, which may be used when the user has a severe or profound hearing loss in one ear and a moderate hearing loss in the other. In this case, two microphones and one amplifier are used, providing the user with an amplified signal from both sides.

Hearing Aid Models

Hearing aids are divided into four main groups.

1. Behind The Ear (BTE)
2. In The Ear (ITE)
3. Body worn aids
4. Hearing glasses

The BTE hearing aid is a widely used model. It is placed behind the ear, and the sound is conducted to the ear canal through a plastic tube and an earmould. BTEs can be used for the majority of hearing losses. They are the typical choice for severe losses due to their gain and output capabilities and the possibility of combining them with educational equipment via audio input.

The ITE hearing aid is placed directly in the cavity of the outer ear. All parts of the hearing aid are built into the earmould. One advantage of the ITE hearing aid is that the “natural” effect of the outer ear (pinna) is maintained, making it easier to determine the direction of the sound source and take advantage of the natural high frequency amplification of the pinna.

ITEs can be modular or custom made. In the *modular* type, the hearing aid is a complete functioning unit so small that it can be placed entirely within or onto an earmould. This makes it easy to replace and repair the hearing aid. *Custom made* models can be made slightly smaller in size, but at the expense of ease of replacement and repair. Both types are available in two basic models: full concha and canal.

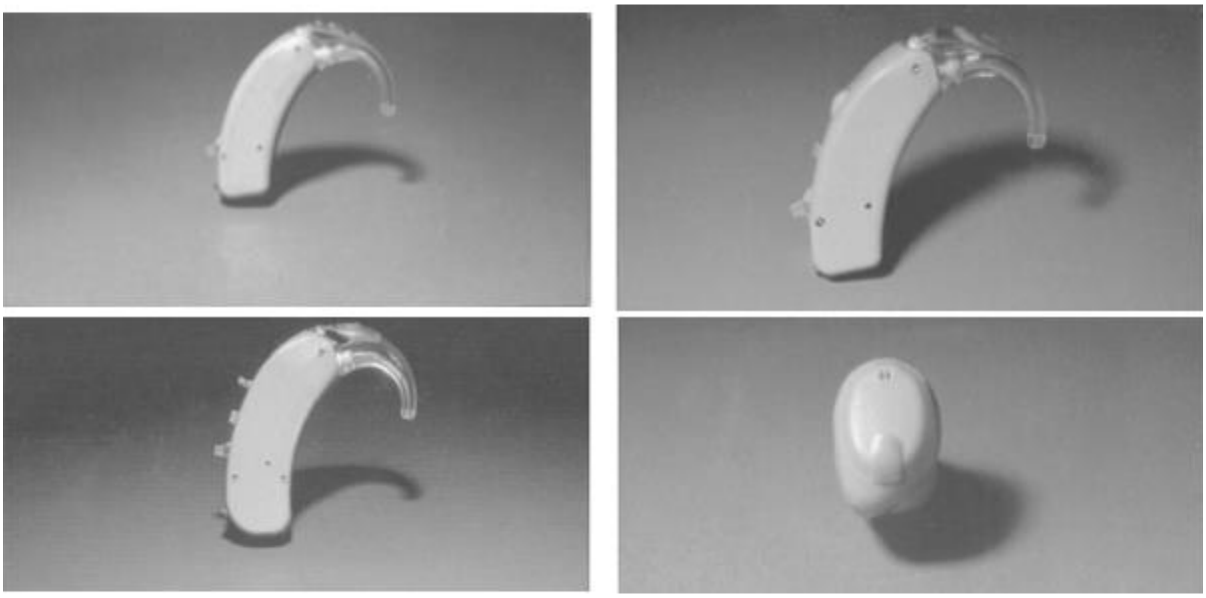


Fig. 3.13 Hearing aid models: Three typical behind the ear (BTE) models and a typical in the ear (ITE) hearing aid.

INSIDE A HEARING AID

The main components of a hearing aid are :

1. An *input transducer*, (a) microphone or (b) telecoil
2. An *amplifier* with volume control, filters, automatic gain control (AGC) and max. output regulation.
3. A *receiver* (loudspeaker)
4. A *battery*

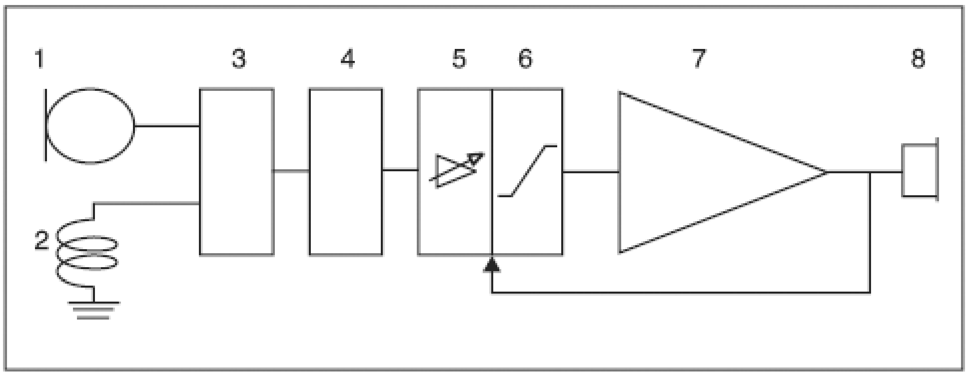


Fig. 3.14 Simplified diagram of a hearing aid: 1. Microphone, 2. Telecoil, 3. Input amplifier, 4. Filters, 5. Volume control, 6. AGC/limiter, 7. Power amplifier, 8. Receiver.

The Microphone

The microphone converts sound waves into electrical signals. The electret microphone is the most commonly used type. It is a special type of condenser microphone with a built-in low-noise preamplifier. The electret microphone has a smooth frequency response and a low sensitivity to mechanical vibrations. This reduces the risk of *acoustic feedback.*

Hearing aid microphones exist in a number of versions having different frequency responses and sensitivities. Some hearing aid types are available with optional microphones. The most commonly used microphones are; normal range, wide range, 6 or 12 dB per octave ski slope, step response and damped peak microphones.

As an alternative to shaping the frequency response with special microphones, electric filtering techniques have become more common. These techniques may be preferable with regard to reducing internal noise from the microphone.

Microphones can be omnidirectional or directional. Most microphones used in hearing aids are omnidirectional, meaning that sound is received from all directions with almost equal sensitivity.

To improve speech discrimination in noise, a directional microphone can be used in BTE hearing aids with the effect that signals coming from behind are being suppressed, thus improving the user's ability to attend to the frontal sound source. In reverberant situations, the positive effect of directional microphones is diminished.

Note that directional microphones normally have reduced sensitivity at low frequencies compared to normal microphones.

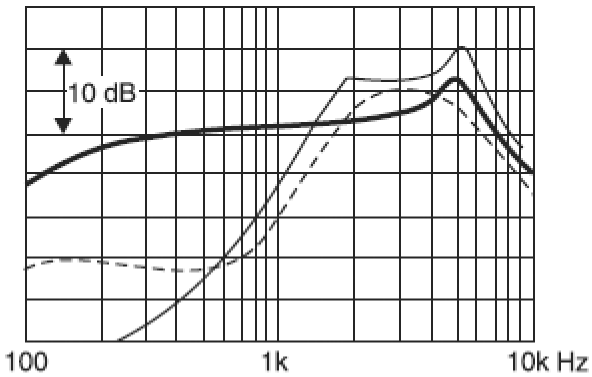


Fig. 3.15 Frequency response of three types of microphones: normal range (Thick line), step response (Dashed line) and 12 dB-per-octave ski slope (Thin line)

The Telecoil

A telecoil transforms variations in a magnetic field into electrical signals which can substitute the signal from the microphone. The telecoil is used in connection with a teleloop system, which are installed in places such as theatres, churches and schools for the hearing impaired.

A *loop system* consists of a loop amplifier and a teleloop. A well functioning system produces a magnetic field representing the sounds to be heard. Hearing aids equipped with a telecoil can convert this magnetic field into sound when the telecoil is activated (M-T switch in position T). In this way, background noise may be reduced. A small telemagnetic system can be used in private homes for listening to TV. A telephone adapter can be used for telephone communication.

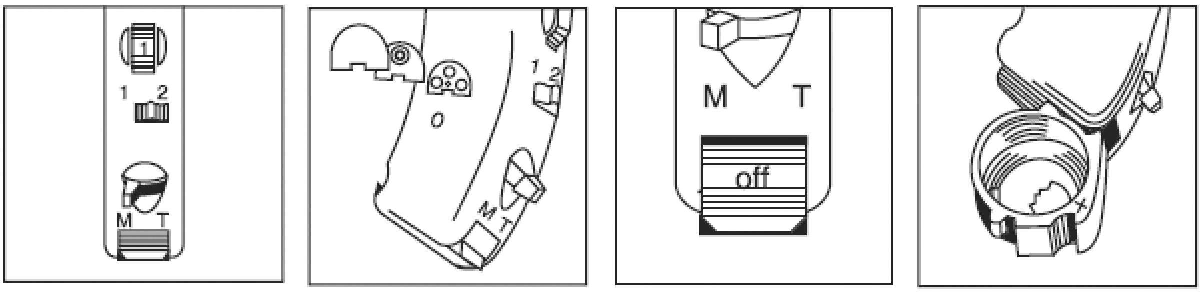


Fig. 3.16 The logical user controls

Audio Input

A good sound quality can also be obtained by using an electrical input connection directly on the hearing aid, known as an *audio input system*. This is often used in educational situations, and frequently in connection with radio (FM) or infrared transmission.

The Amplifier

Signals from the microphone or the telecoil are fed to the amplifier. *This weak low-energy signal is transformed into a powerful electrical signal which drives the hearing aid receiver*. There are different types of amplifiers. The *class A amplifier* uses the same high amount of current whether there is a signal at the microphone or not. It has now been almost completely replaced by amplifiers with lower current consumption.

The linear push-pull amplifier *(Class* B), and modern variations of it, are far more economical and have very little distortion. Power consumption is low whenever the signal is weak, and higher when a stronger signal is present at the microphone.

Switching amplifiers utilise pulse modulation to increase the useful effect of the amplifier. The amplifier may be partly integrated into the receiver. Switching amplifiers are characterised by low distortion and high efficiency. One type of switching amplifier is called *Class D.*

Amplifiers can be made either with *discrete components* mounted onto a printed circuit board, as thin-film or thick-film circuit blocks, as *integrated circuits* (IC), or as a *combination* of these.

The Discrete Amplifier

The discrete amplifier has *individually mounted components* and it takes up more physical space than other types of amplifiers. It is, therefore, used primarily in large hearing aids such as body worn aids and large BTEs.

Thin Film Amplifier

The thin film amplifier is built onto a thin ceramic or glass plate covered with a resistive material. Individual resistors are made by an etching process and the connections are achieved by adding gold to the surface. By using *thin film techniques*, it is possible to make very precise resistors. Transistors, integrated circuits and capacitors are mounted onto gold connectors. Connections are made with thin gold wires, using a special bonding process.

Thin film circuits normally have only one layer of connections on each side. The connections from one side to the other are difficult to make and are relatively large and thereby impractical. Consequently, thin film circuits are often made *single sided.*

Thick Film Amplifier

The thick film amplifier consists of a thin ceramic plate, on which resistors and connectors are "printed" onto the surface by using a special *silk screen technique.* These resistors and connectors are larger than those for thin film, and the resistors have to be adjusted if they are to be precise. It is possible to make many layers and to make connections between the layers and the two sides. *The thick film technique is the most commonly used process today.* It is often used in connection with integrated circuits. Transistors, integrated circuits and capacitors are mounted onto the surface, often on both sides of the ceramic plate.

Integrated Circuits (IC)

The IC component, the size of which is often only a few square millimeters, may contain several thousand transistors and resistors. *The IC has made it possible to make hearing aids smaller and more sophisticated than before. Hearing aids often use an IC which is "custom" made for a specific hearing aid model.*

In the process of designing an IC, Computer Aided Design tools (CAD) are used. Generally, the design phase has become more complex, while the production phase has been simplified.

The Receiver (Loudspeaker)

The receiver converts the amplified electrical signal into sound waves. There are different receiver types, both in size and performance. *Generally, size determines the sensitivity and the maximum output of the receiver.* By using an acoustic filter, the frequency response and the frequency range may be improved.

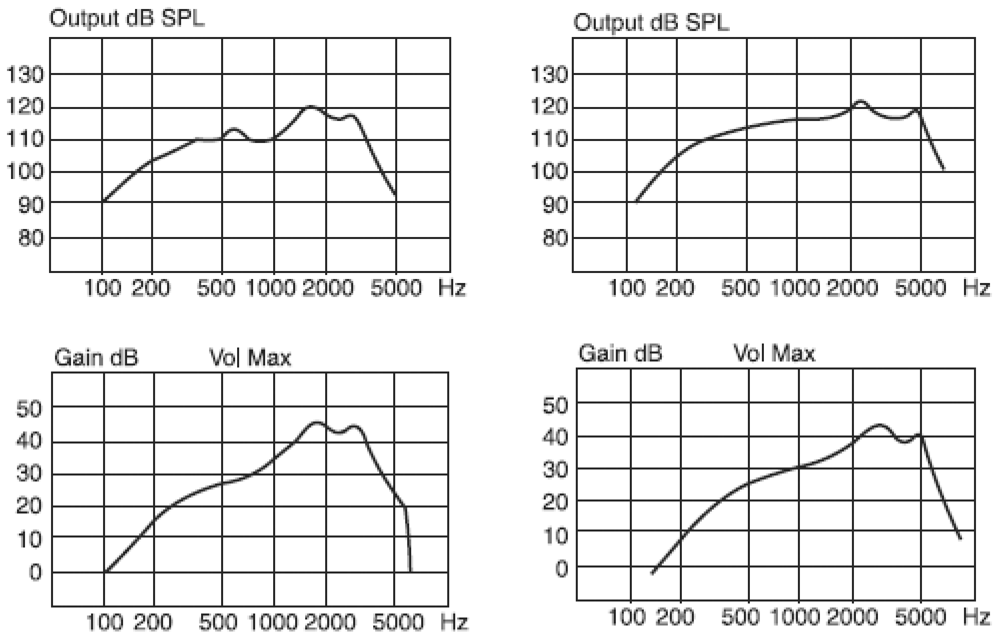


Fig. 3.17 Maximum output and frequency response curve for a hearing aid with normal (left) and wide range (right) receivers.

Batteries

Batteries are available in different sizes and types and the most commonly used are *zinc air batteries*. Zinc air batteries have high energy density, they are affordable, and used batteries have little impact on the environment.

Battery capacity is measured in milli-ampere hours (mAh). For instance, if a battery having a 100 mAh capacity is used in a hearing aid with a battery drain of 1 mA, it will last approximately 100 hours. *Battery drain* for modern amplifiers depends on the output produced by the hearing aid. Therefore, battery drain cannot always be specified exactly.

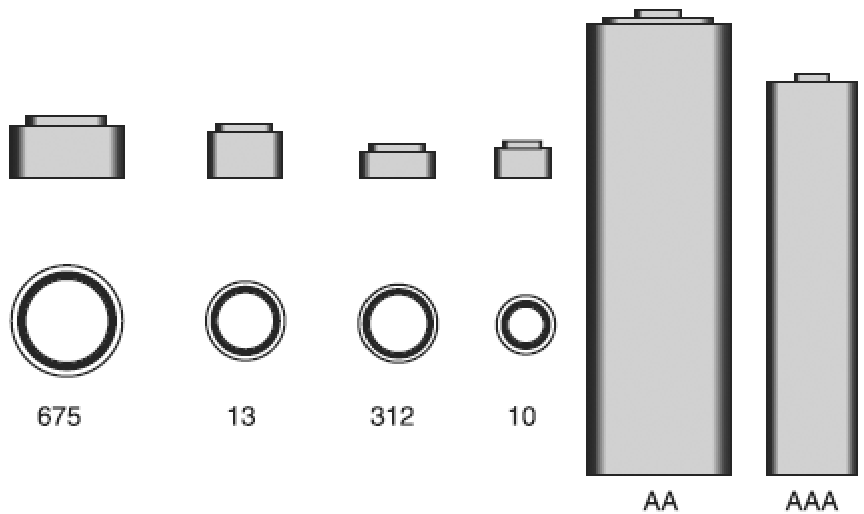


Fig. 3.18 The most common battery sizes : Type 675 (11.6 x 5.4 mm), type 13 (7.9 x 5.4 mm), Type 312 (7.9 x 3.6 mm), type 10 (5.9 x 3.6 mm) and Penlight batteries type AA (14.5 x 50.5 mm) and AAA (10.5 x 44.5 mm).

For hearing aids with extremely high power consumption, or for powerful aids in heavy duty use, *mercury or silver oxide batteries* are still used. However, modern zinc air batteries are very competitive, coming close to the performance of mercury and silver oxide batteries.

Alkaline batteries are used for body worn instruments and hearing aid remote controls.

USER OPERATED CONTROLS

The Volume Control (VC)

Most hearing aids have a volume-control, which enables the user to adjust the amplification of the hearing aid according to the *different listening environments.* Some hearing aids have no VC accessible to the user. Instead they are equipped with some sort of *automatic gain control* and the gain is preset by the manufacturer or the hearing healthcare professional.

The On/Off switch

The hearing aid is turned on or off by using the on/off switch. This can either be a tiny switch, mounted on the hearing aid case, or it can be *integrated* into the battery compartment for ease of operation.

The M-T switch

Many hearing aid models are provided with a switch, enabling the user to choose between receiving the signal, either from the *microphone*, or a *telecoil loop system.* This is called the M-T switch. The hearing aid receives the signal from the telemagnetic loop system whenever the M-T switch is in the T position. The telecoil system is often used in theatres, cinemas and churches. With a well adjusted telemagnetic system, the use of the telecoil, instead of the normal microphone input, may improve sound quality significantly in situations where background noise is present and there is a relatively large distance to the sound source.

When the hearing aid is in T mode, the microphone is disabled. In some situations, it may be advantageous to use T and M simultaneously, e.g. when listening to the TV via the telecoil, while still wanting to hear speech, the door bell or other important environmental sound. For this purpose, some hearing aids have an MT mode which activates both the microphone and the telecoil simultaneously.

The N/H switch

Some hearing aids have a *normal response/high tone switch* (N-H switch), often instead of the M-T switch. When placed in the H position, the N-H switch will change the sound from normal response to high tone emphasis by attenuating the low frequencies. This improves speech discrimination in noisy environments, e.g. traffic, subways, in a car.

Program selection switch

Some multi-program hearing aids have a *program selection switch* on the hearing aid.

Remote control

Many hearing aid users prefer to control their hearing aids in a more sophisticated manner. They want to select from a larger number of features, depending on the *listening situation.* Since the hearing aid user often wants a rather small and cosmetically acceptable hearing aid, such controls cannot be implemented in traditional ways with switches and trimmers.

Instead, it is possible to provide hearing aid models with several *user selected programs.* Such a program contains a combination of acoustic parameter settings, matched to provide optimum listening for a particular user in a certain listening situation. By using a *remote control*, the user can select from a number of preset programs. The transmission pathway from the remote control to the hearing aid is either based on infrared light, ultrasonic sound, or radio waves. *By the simple press of a key on the remote control unit, the hearing aid user has immediate access to the best acoustic response for a given listening environment.*

DISPENSER OPERATED CONTROLS

Most hearing aid models have facilities to adjust the hearing aid for the *individual hearing loss.* These facilities, which vary from switches and trimmers, to computer operated digital memories, are used to **adjust** e.g. amplification of low or high tones, maximum output, preset gain and compression (automatic gain control (AGC)).

Preset Gain Control

The hearing aid often has a preset gain control permitting a *reduction* of the maximum gain by up to 20 dB. The preset gain is adjusted to a position giving the user the most suitable operating range for the volume control. The *preset gain control* is also used to reduce maximum gain in order to avoid *feedback* (whistling). This function can be useful for children or for those having difficulties adjusting the volume control.

Preset gain control can also be used to create binaural loudness balance, as well as to equalise the balance for different programs in a multi-program hearing aid.

FILTERS

Low-Cut Filter

The low-cut or *bass-cut filter* is used to reduce amplification of low frequencies. This filter is used if the hearing impaired person has normal or near normal hearing in the low frequencies, or experiences trouble hearing in noisy environments.

There are various degrees of filter slopes, ranging from a slope of 6 dB per octave in a traditional, *first order passive filter,* to 12, 18 or 24 dB per octave in an *active filter* (which contains an amplifier). Special tone filters may have a **variable slope.**

The number of filtering components (such as capacitors) determines the order of the filter. A first order filter can be designed using only one capacitor, while four are needed in a 4th order filter.

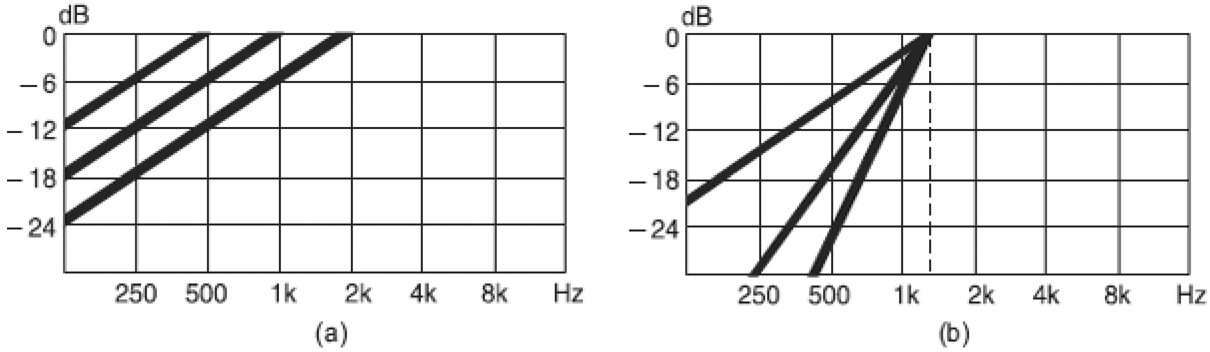


Fig. 3.19 Low cut filters. (a) With different cut off frequencies. (b) With different slopes

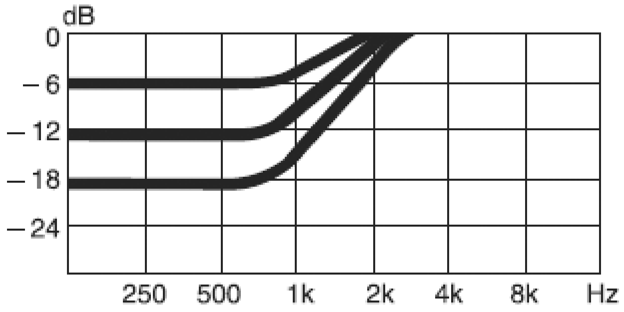


Fig. 3.20 Frequency response of the inverse presbycusis adaptation (IPA) filter.

The majority of hearing impaired people suffer from hearing loss due to the aging process (presbycusis). They hear low frequencies without significant problems, but have impaired hearing in the high frequencies. Therefore, a special filter, the *Inverse Presbycusis Adaptation (IPA) filter*, has been designed to compensate for this type of loss.

High-Cut Filter

A high-cut filter is used to reduce amplification of high frequencies, normally above 1000 Hz. This reduction is also useful for those with a severe hearing loss, to reduce the risk of acoustic feedback. The filter may, furthermore, be used for first time hearing aid users, *in order to attenuate the high frequencies*. This is often preferred in the early stages of hearing aid use. After a while, the user becomes accustomed to the hearing aids. It is then possible to change the high-cut filter response to increase amplification of the high frequencies.

The *high-cut filter* can also be a first order or higher order filter, as described for the low-cut filter.

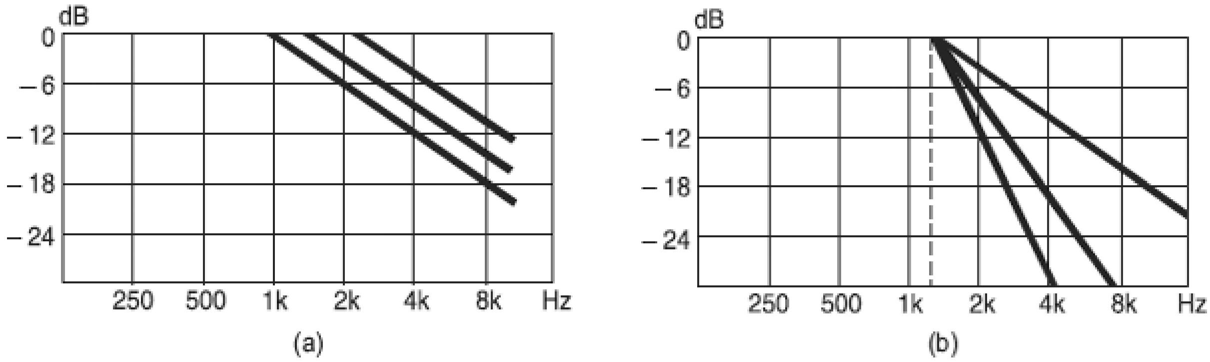


Fig. 3.21 Frequency response of a high cut filter (a) adjusting the cut-off frequency (b) adjusting the slope.

OTHER TYPES OF AGC

A large number of AGC types have been developed during the past decades. They all have the common purpose of automatically adjusting the gain and/or the frequency response of the hearing aid in accordance with the input signal and the user's dynamic range.

Multi-Channel AGC

In multi-channel AGC, the input sound is divided into two or more frequency channels, each of which has its own AGC circuit. Multi-channel AGC has the positive feature that a loud sound in one frequency band will not affect amplification in a different frequency band. On the other hand, there is a tendency for a spectral smearing diminishing spectral contrasts. Such smearing may have a negative effect on speech intelligibility, and hence hearing aid acceptance.

Automatic Signal Processing, ASP

ASP essentially denotes *a way of automatically adjusting the low frequency cut-off* according to the content of intense, low frequency components of the input sound. If dominating low frequency components are present in the input signal, as e.g. in a car or in heavy traffic, the low frequency gain is automatically reduced to prevent spread of masking.

Such a **bass reduction** may also prevent the hearing aid from being overloaded and sounding distorted. **Loudness Compensating AGC**

A number of AGC circuits, including the so-called K-amp, aim at **compensating for abnormal growth of loudness**, which is seen in many hearing impaired persons. The net effect is that the need for adjusting the volume control becomes less.

Adjustment of hearing aids with loudness compensating AGC is often quite specific for each type of instrument. In general, the adjustment is performed on the basis of the audiogram, as well as measurements of loudness perception. Usually the fitting is carried out using a computer or a dedicated programming device.

Descriptive Questions

1. Differentiate between headphones and headsets.
2. What are the different types of headphones based on their construction?
3. Explain the working of a dynamic headphone.
4. Discuss the working of electrostatic headphones. How do they differ from other headphones?
5. What are the peculiarities of electret-electrostatic headphones?
6. Discuss the limitations of various types of headphones?
7. What are the requirements of a hearing aid?
8. Explain the different types of hearing aids.
9. What are the controls associated with hearing aids?

Multiple Choice Questions

1. Moving iron headphones work on the principle of
    a. magnetic attraction
    b. magnetic repulsion
    c. both (a) and (b)
2. Moving iron headphones suffer from
    a. natural resonance
    b. high harmonic distortion
    c. non-linear distortion
    d. none of these
3. The drivers of dynamic headphones can be expected to produce
    a. ample sound pressure level
    b. insufficient sound pressure level
4. Electrostatic headphones
    a. suffer from system resonance
    b. are free from system resonance
5. Which one of the following suffers from high harmonic distortion
    a. moving iron headphones
    b. moving coil headphones
    c. electrostatic headphones
    d. crystal headphones
6. Which one of the following headphones needs polarizing voltage

    a. moving coil
    b. electrostatic
    c. moving iron
    d. crystal

Fill in the Blanks

1. _____________ permit one-way communication.
2. _____________ permit two-way communication.

3. Interaction between the two fields in moving iron headphones varies the _____________ on the diaphragm causing it to _____________.
4. Plastic membranes can have much less _____________ and _____________ than metal diaphragms thereby minimising problems with _____________.
5. The impedance of crystal headphones is predominantly_____________.
6. Electrostatic headphone lends itself to a _____________ configuration minimising _____________.
7. Electret electrostatic headphones are permanently _____________.
8. LED indicators in headphones indicate that the circuit is _____________.

ANSWERS

Multiple Choice Questions

1.	(a)
2.	(a)
3.	(a)
4.	(b)
5.	(a)
6.	(a)

Fill in the Blanks

1.	Headphones
2.	Headsets
3.	pull, vibrate
4.	mass, stiffness, system resonance
5.	capacitive
6.	balanced configuration, harmonic distortion
7.	polarised
8.	activated

CHAPTER 4

LOUDSPEAKERS

We have progressed a long way since early designers fixed a crude horn to a telephone earpiece to make the first loudspeaking telephone. Since then, many other drive principles have been adopted to further the quest for perfect reproduction. The science of loudspeakers has been refined to the point where the response of a drive unit, at least at low frequencies, can be predicted with great accuracy. Just as the choice of a microphone dictates what goes into the recording, the loudspeaker decides the quality of the output.

IDEAL LOUDSPEAKER

There are a number of *interrelated factors* that must be considered in designing transducer for converting *electrical energy* into airborne *acoustic energy*. These include electroacoustic efficiency, uniformity of frequency response, linearity of amplitude response, transient response, power handling capacity, size, durability and cost. An *ideal loudspeaker:*

1. would have an *electroacoustic efficiency* approaching 100 per cent.
2. would have an *acoustic output response* that is independent of frequency over the entire audible range.
3. would introduce neither harmonic nor intermodulation *distortion* into its output.
4. would *faithfully reproduce* transients as well as steady input signals.
5. would be capable of producing a *nondirectional radiation pattern.*
6. would be of *as small a size as is possible* considering the required acoustic output.

No single transducer has been designed that is capable of satisfying all the above requirements. Out of many devices developed for the radiation of acoustic energy into air, the two most widely used are the *direct-radiator* or *dynamic loudspeaker* and the *horn loudspeaker.* Both of these loudspeakers utilise the *electrodynamic coupling* that exists between the motion of a vibrating surface, called the *cone* or *diaphragm* and the current in a so called *voice-coil.* Additional types of *electromechanical coupling* that are used for this purpose include *electrostatic coupling* in electrostatic loudspeakers and *electromagnetic coupling* in telephone receivers.

The speaker system itself can be divided into three *functional parts*:

1. *The electromagnetic part*, consisting of the voice coil and the field magnet. Audio frequency electric current in the coil causes mechanical motion of the cone or diaphragm on which it is mounted. This part is often referred to as the *driver or motor* of the system.
2. *The mechanical part,* on which the driving coil is usually mounted and which is set into mechanical motion by the audio frequency electric current in the *driving coil.*
3. *The acoustic part*, which transmits the sound energy developed by the mechanical part of the area served by the system in the most efficient and faithful manner possible. This takes the form of a *baffle* or *enclosure* with a horn being a form of enclosure.

A complete understanding of the operation of the speaker systems requires a sufficient view of the *flow* of acoustic energy from the ouput amplifier stage to the listener.

BASIC LOUDSPEAKER

When broadcasting first began, the only available devices for converting sound energy into electrical energy, and vice versa, were those used in the ordinary telephone. Microphones were usually of the carbon type. The hearing device, or telephone receiver, was a simple electromagnet energised by electrical signals. These signals caused a thin sheet of magnetic material (the telephone diaphragm) to vibrate in *synchronism.* Wearing headphones, however, was not exactly conducive to listening pleasure, besides the reproduction wasn't of the best quality. So, engineers started to search for *some sort of hearing device that could be separated from the head of the listener and still supply enough volume to make listening easy.* Hi-fi wasn't really thought of at that time.

The two problems faced by the early engineers were; *amplifying* the electrical signal enough to drive the loudspeaker, and designing an *off-the-head speaker* that could reproduce the electrical signals in *the form of sound.*

One of the first speakers designed was simply a large *horn* attached to an ordinary telephone earphone mechanism. This speaker had fairly good volume, but frequency response was poor.

In another early design, a *vibrating reed* was used instead of the thin disk of magnetic material (the diaphragm of the earphone). The free end of the reed was attached to a speaker cone, and the reed's vibration caused the whole cone to vibrate.

A third type of speaker had a *balanced pivoted armature* located between a pair of magnets. Electrical impulses through the magnet coils caused up-and-down vibration. The armature movement was mechanically coupled to the speaker cone with a stylus, thus producing the sound.

There are three types of speakers in modern use: crystal, electrostatic (condenser) and electrodynamic (or simply dynamic) speakers. Dynamic loudspeakers are illustrated in Fig. 4.1.

Fig. 4.1 Dynamic loudspeakers

CRYSTAL LOUDSPEAKERS

Rochelle-salt crystals have the property of becoming *physically distorted* when a voltage is applied across two of their surfaces. This property is the basis of the crystal type of speaker driver, illustrated in Fig. 4.2.

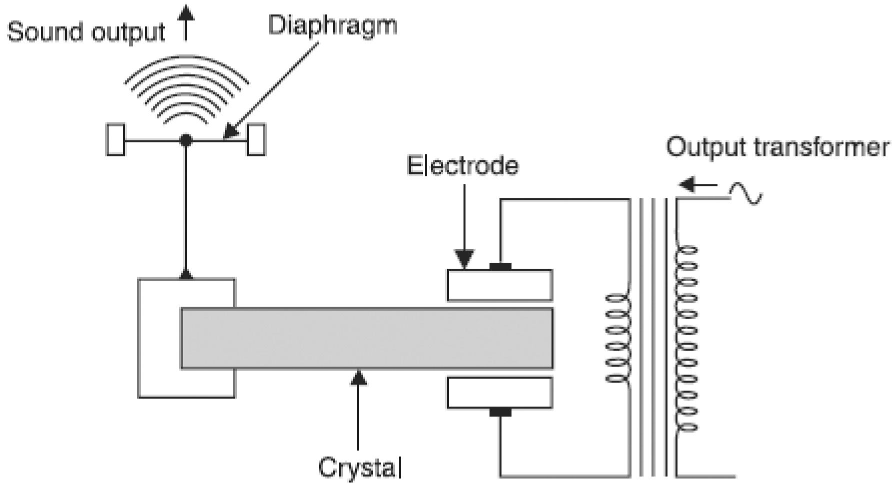


Fig. 4.2 Schematic representation for a crystal type speaker

The crystal is clamped between two electrodes across which the audio frequency output voltage is applied. The crystal is also mechanically connected to a diaphragm. The deformations of the crystal caused by the audio frequency signal across the electrodes cause the diaphragm to vibrate and thus to produce sound output.

Crystal speakers have been impractical for reproduction of the full audio-frequency range because *the input impedance is almost completely capacitive.* Thus it is difficult to couple power into them. At high audio frequencies, the reactance becomes lower ($X_c = 1/2\pi f C$) and the relative amount of power smaller. In the bass range, stresses on the crystals are very great, and crystals have been known to crack under stresses.

Consequently, crystal units have found some use in *tweeters* (the high-frequency portion of dualspeaker units) and rarely even in this application because *their response is not linear.*

ELECTROSTATIC (CONDENSER/CAPACITOR) LOUDSPEAKERS

This type of speaker operates on the principle that a dc voltage between two parallel metal plates causes these plates to attract or repel each other. The amount of attraction or repulsion depends on the applied voltage. If one of the plates is a flexible metal, it will bend. But *the amount of attraction and repulsion is not directly proportional to the voltage applied.*

For example, consider the movable and fixed plates of Fig. 4.3 with *no voltage* applied. Now suppose we apply a *slowly varying ac voltage* to both plates. As the voltage increases from zero, the potential difference between the two plates also increases. This, in turn, produces an increasing force of attraction between the plates, so that the movable plate bends towards the fixed plate. As the ac voltage decreases once more to zero, the attractive force decreases, and the movable plate moves back to its original position. But, now we have the second half of the ac cycle, in the *negative direction.* All that this means to the metal plates is that the positive and negative voltages have **switched** plates. The attractive force is still there, and it is still the same. So, we get another bend in the movable plate on the negative half of the ac cycle. Thus, for one full cycle of ac we have two bends in the movable plate, in effect a *frequency doubling.* A 2 kHz signal would give us a 4 kHz note.

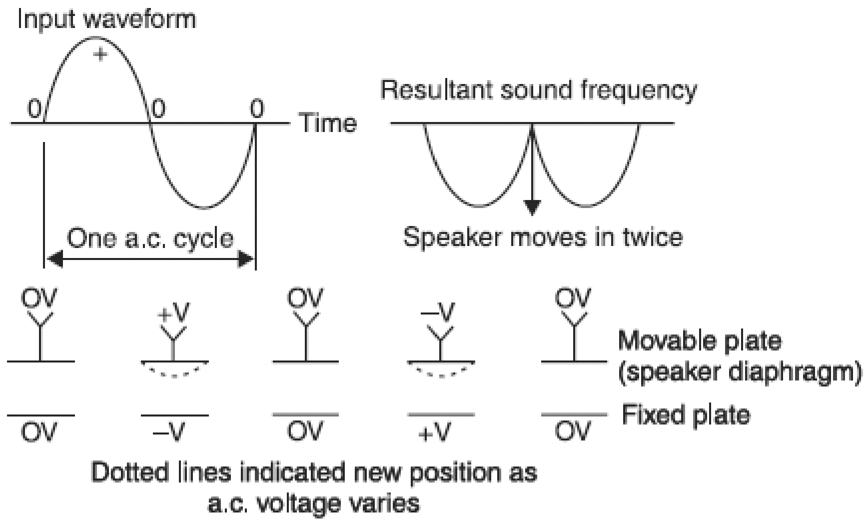


Fig. 4.3 Frequency doubling in an unpolarised loudspeaker

To overcome frequency doubling, we polarise the speaker, that is, we apply a high voltage (1,000 volts or so) as a sort of dc *bias*, (Fig. 4.4). The voltage exerts a *steady attraction* between the two plates, so that now—with no signal—*the movable plate is bent slightly toward the fixed plate.*

Now suppose we apply a 400 V dc audio signal to the speaker. As the positive half cycle of the signal increases from zero the voltage between the plates *rises* from 1,000 V toward 1,400 V and the movable plate bends from its original position toward the fixed plate. As the ac passes its peak and returns to zero, the voltage between the plates *drops* from 1,000 V to 600 V. Instead of moving again toward the fixed plate, the movable plate moves farther away. *So we have a situation in which the bending of the movable plate is identical to the ac swing and there is no frequency doubling.*

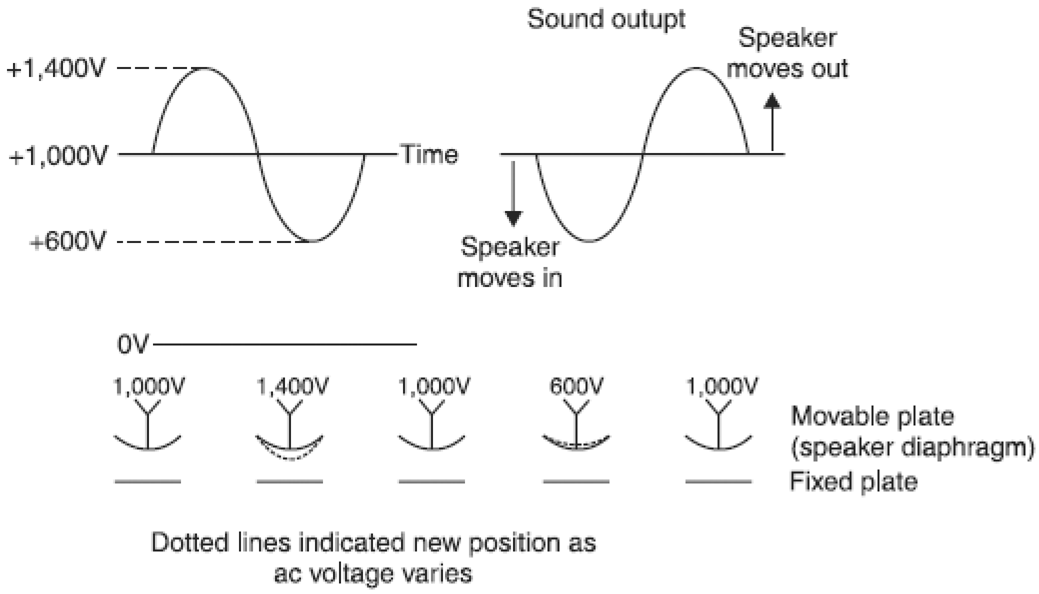


Fig. 4.4 Frequency doubling eliminated by dc polarisation

A detailed view of a modern electrostatic speaker is shown in Fig. 4.5. The practical speaker of today uses push-pull, with a built-in step-up transformer to work from the ordinary 8 ohm amplifier output tap. The polarising voltage is applied to the centre or movable plate through a resistor that keeps the voltage *stable* during variations in the signal voltage. The signal voltage is applied to the two outside plates. Because the diaphragm is centered between the two plates that attract it equally, there is no bending when there is no signal. Also, *because of the push-pull action the diaphragm can move twice as far in response to signal voltages for the same amount of compression of the dielectric material.*

The major weakness of the electrostatic speaker requires the dc bias is that it to be much larger than the applied audio signal. In practical speakers, 1,000 to 1,200 volts may be used. Further, when we get into the bass frequency ranges, a great deal of power would be required to get enough output. To produce such power, the speaker area would have to be very large. So, even though full range electrostatic speakers have been constructed, *in practical use electrostatic speakers have been mostly confined to frequencies above 1,000 Hz.*

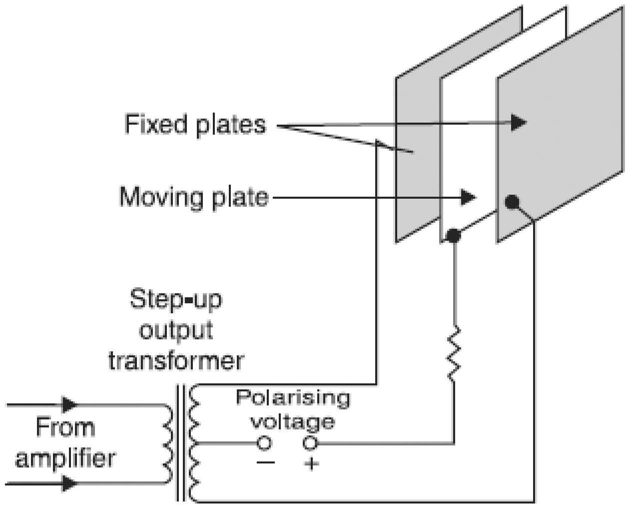


Fig. 4.5 Working principle of a typical electrostatic loudspeaker

The step-up transformer and the high voltage polarising supply is usually built right into the modern electrostatic. Often the electrostatic unit and its matching woofer are sold together as a complete system.

Some high class systems use electrostatics to reproduce the high frequencies. Koss uses electrostatics on some of their stereo headphones.

DYNAMIC LOUDSPEAKERS

There are two varieties of dynamic loudspeakers : electrodynamic and permanent magnet (PM) speakers. Both work in exactly the same way, the difference is in their construction.

The *electrodynamic* speaker has a *soft iron* magnetic circuit, non-retentive of magnetism, around whose centre leg, a large, multilayer field coil is wound, as shown in Fig. 4.7. When dc flows through this field coil, it magnetises the iron core. *A magnetic flux field directly proportional to the strength of the current through the coil is thus set up across the airgap. The iron core is not permanently magnetised, it stays magnetised only as long as current flows through the field coil.*

Fig. 4.6 A two-way electrostatic utilising a separate woofer and tweeter.

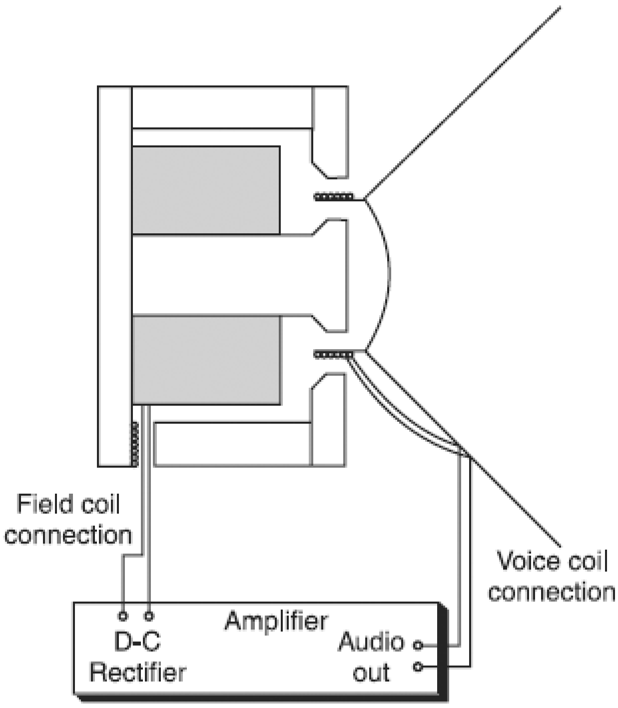


Fig 4.7 Electrodynamic speaker

Improvements in permanent magnet materials have made the electrodynamic speaker *practically obsolete*, but some still exist in vintage radios. Note that these use the field coil as part of a choke filter in the power supply, a good example of killing two birds with one stone. The electrodynamic speaker has disappeared completely, so far as hi-fi is concerned, the permanent magnet speaker reigns supreme.

PERMANENT MAGNET LOUDSPEAKERS

The *most popular* type of loudspeaker today is the permanent magnet dynamic type. Because of its comparative **simplicity** of construction and design, the **precision** that may be built into it, the **ease** with which it is interfaced with other equipment, its easy **adaptability** to many different applications, and its **comparative freedom** from electrical trouble, **the *dynamic loudspeaker* has found acceptance in all kinds of reproducing systems.** It is found in the smallest pocket radios and is a major component of the most elaborate theatre systems (See Fig. 4.8).

Just about all hi-fi woofers are of the permanent magnet (PM) type. Exploded view of the PM cone type speaker is given in Fig. 4.9. The *cone* (diaphragm) is energised by a *moving coil.* The woofer's magnetic field is supplied by a *permanently magnetised and highly magnetic alloy* instead of the iron-cored coil used in electrodynamic speakers.

The PM speaker contains a very light coil of wire affixed to the diaphragm and located concentrically around, within, or in front of the centre of the permanent magnet. The coil (voice coil) is free to move in the field of the magnet. Electrical impulses, varying at an audio rate, are applied to the *voice coil* by the amplifier. Because these impulses are constantly changing in amplitude and direction, a changing magnetic field is set up in the voice coil. This field *reacts* with the constant field of the permanent magnet. The result is that the voice coil moves further into the gap when the fields are opposite and attract, and farther out of the gap when they are alike and repel. This causes an *in-and-out movement of the diaphragm*; consequently, we obtain sound waves from electrical impulses. The speed at which the coil and diaphragm vibrate depends upon the *frequency* of the impulses. The distance that the diaphragm moves in and out depends on their *amplitude.*

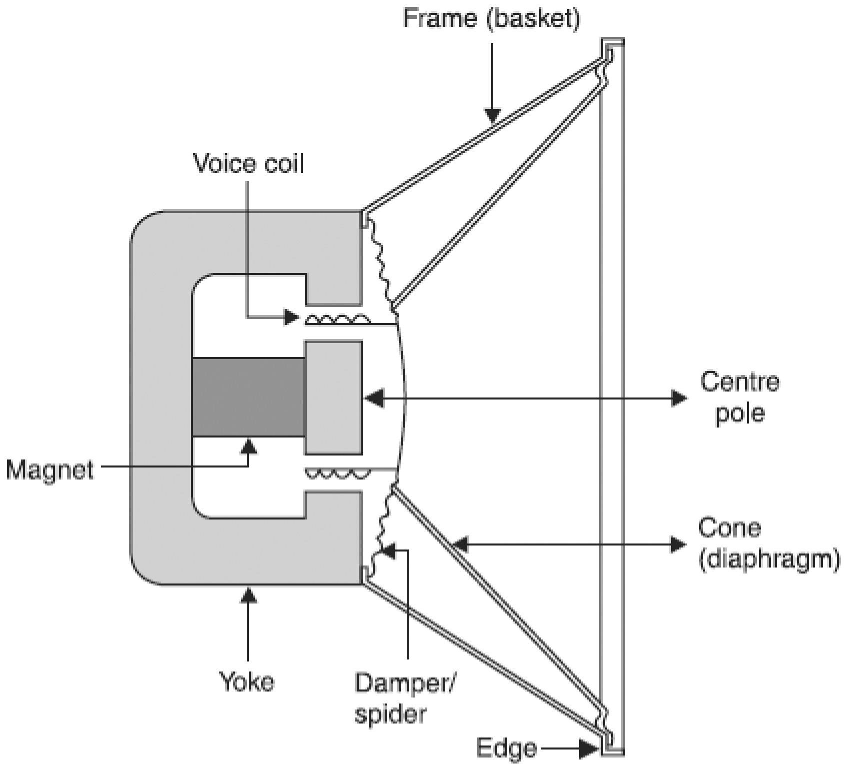


Fig. 4.8 Dynamic loudspeaker

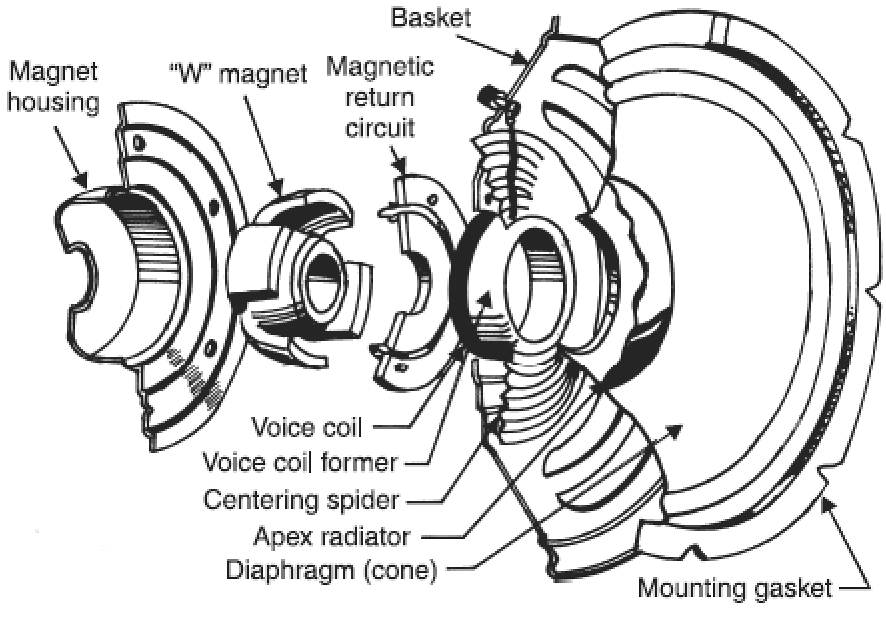


Fig. 4.9 Exploded view of dynamic loudspeaker

LOUDSPEAKER CONSTRUCTION

The *voice-coil* is wound on a *cylindrical form*. The *diaphragm*, usually made from a special paper, is attached to the *outer rim* of the voice coil form. Some speaker systems utilise speakers that have aluminium diaphragms. The metal prevents the effects of humidity changes and helps dissipate the thermal energy present when the speakers are driven at high levels by powerful amplifiers. High quality speakers use diaphragms composed of titanium, aluminium, and paper to achieve the *maximum stiffness* required for their mode of operation.

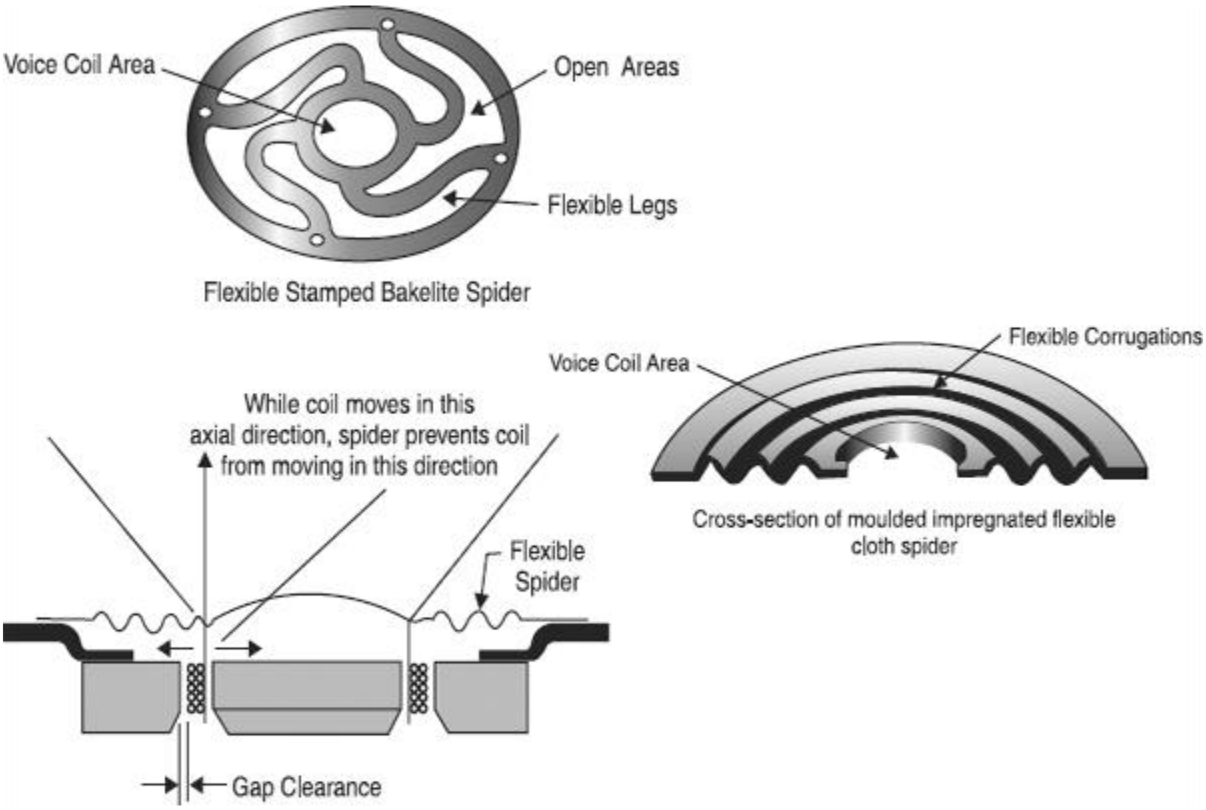


Fig. 4.10 Some typical voice coil centring devices

The outer edge of the diaphragm is cemented to the speaker's metal frame (basket). The permanent magnet is located concentrically in the back of the voice coil. To prevent the voice coil from shifting, it must be supported adequately and maintained at *dead centre;* however, it must be free to move in and out. This is accomplished by the flexible *spider*, Fig. 4.10. With the latter cemented firmly in place, *the voice coil is free to move in and out, but not vertically or laterally*.

PERMANENT MAGNET

The strength of the permanent magnet largely determines both speaker efficiency and quality of reproduction. The magnet must furnish the most powerful magnetic field in order to maximise the voice coil movement for any given signal. In general, *the stronger the magnet, the better the power handling capacity of the speaker*, *and the better the reproduction.*

Woofers are speakers designed to reproduce the bass, or *low frequency,* portion of the audio-frequency spectrum. For woofers, the designer does not choose the strongest available magnet as *adding more magnetic flux will reduce the bass performance of a woofer because the moving system becomes overdamped.* The designer, therefore, chooses the magnetic flux level that provides the best compromise between *efficiency* and *bass response.* For a hi-fi speaker system this is very seldom the level at which the magnetic flux is as strong as possible.

Fig. 4.11 Typical speaker systems, two-way and three-way

Mid-range speakers are designed to reproduce the mid-frequency portion of the audio frequency spectrum. *Tweeters* are speakers designed to reproduce the treble, or *high frequency* portion of the audio frequency spectrum. As far as the magnetic flux in mid-ranges and tweeters is concerned, the designer usually does try for the strongest flux field possible, of course, within the cost restrictions. A typical speaker system is shown in Fig. 4.11.

The mid-range speaker in a three-way system is often referred to as a *squawker.*

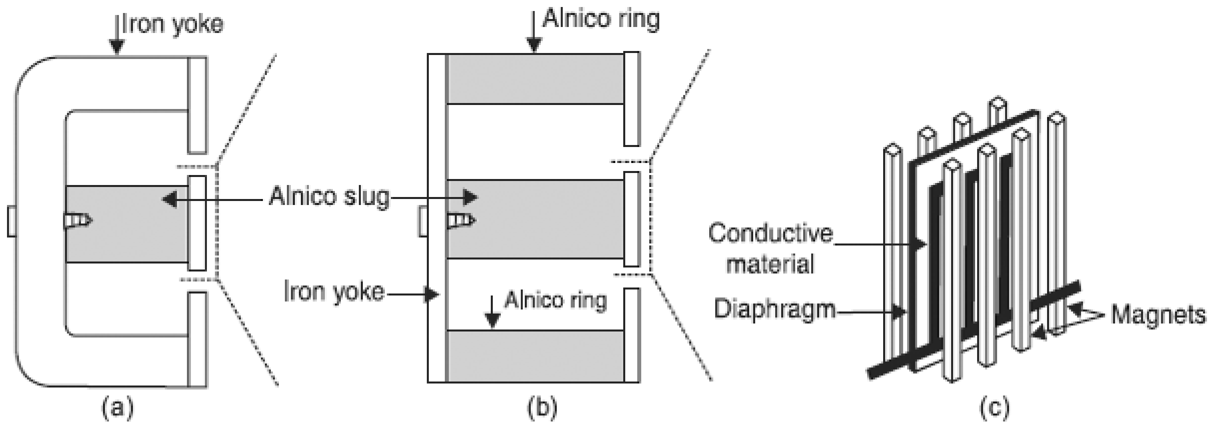


Fig. 4.12 Shapes of magnets in PM speakers : (a) slug-type (b) ring-type and (c) spaced-bar type

Permanent magnets in speakers are available in various *shapes.* One example is the magnet in the form of a *solid cylindrical slug,* as shown in Fig. 4.12 (a). A U-shaped iron yoke completes the magnetic circuit, except for the voice coil gap. Another type, similar in outward appearance, uses a *hollow cylindrical slug* with an E-shaped yoke, with almost the entire magnetic circuit composed of Alnico [Fig. 4.12 (b)]. A third type, Fig. 4.12 (c), uses *spaced permanent magnets.* Here, instead of a voice coil, the diaphragm itself (which consists of Mylar stretched taut over a frame that has copper wire glued in a square-wave pattern to its surface) is placed in the magnetic field.

Ceramic magnets have several advantages over metal magnets. They are lighter, stronger, and less expensive to produce than metal magnets of comparable size. Their development has resulted from the need for *lighter* and *smaller* magnets, but *ones that have strong fields.* Their principle of construction is simple. During the period when the ceramic is moulded to its final shape with the elimanation of impurities, the ceramic is subjected to a very strong magnetic field. Once, the ceramic is cooled, this magnetic field is *retained,* and we have a very strong permanent magnet.

VOICE COIL

The voice coil is the only thing present in the speaker which carries electrical current or signal. It is energised directly from the amplifier. The *voice coil, as its name implies, is the part of the speaker that does the talking.*

The voice coil, Fig. 4.13, consists of several turns of wire wound on a supporting bobbin. Depending on the functional design of the loudspeaker, the *coil* itself may be copper or aluminium wire, although insulated aluminium ribbon is also used. In case of the latter, the ribbon is wound on edge with the flat surfaces of neighbouring turns adjacent to each other and all the turns held together by a binding cement.

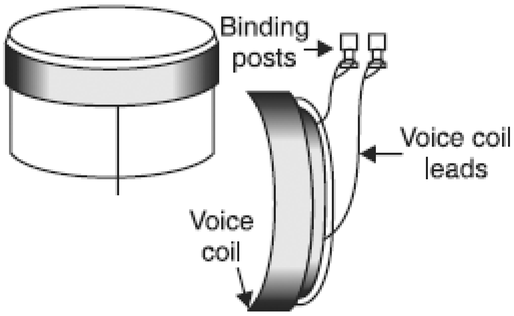


Fig. 4.13 The voice coil

The bobbin, or the *voice coil former* upon which the wire is wound, may be made of a strong grade of thin paper, **wound around on itself several times** to provide a rigid cylinder. Sometimes the voice coil is wound on aluminium or duraluminium formers, and in some designs the voice coil formers are made of rigid paper, reinforced by an aluminium ring around the outer edge. In case the voice coil wrap get twisted, it might contact the core of the magnet. When this occurs the rasping noise develops, rendering the speaker useless. To offset this possibility, many speakers now contain *wrap resistant aluminium base voice coils.*

The voice coil of a dynamic loudspeaker can speak only when it is immersed in a magnetic field. Such a magnetic field may be produced by constructing an *iron loop* with the magnet in one section and an *air-gap* in another, as shown in the magnetic circuit in Fig. 4.14.

The magnet is never charged unless it is in its completed mechanical structure. As soon as the magnet is *charged*, it sets up a complete magnetic field. When electrical signal current from the amplifier flows through the voice coil, we have a *varying* magnetic field of the voice coil in close proximity to the *fixed* magnetic field of the permanent magnet. Motor action will thus be developed. Because of the **interaction** of the two fields the voice coil moves one way or the other, depending on the direction of the signal current as well as the direction of the magnetic field due to the permanent magnet.

The voice coil is attached to the diaphragm which actually fans the air into motion. In order to allow the diaphragm to vibrate back and forth freely, it is necessary to provide it with some sort of flexible support that will allow it to have motion yet keep it vibrating in a true axial direction. The diaphragm is provided with a flexible area at its outer edge *sufficiently compliant* to allow the diaphragm to flex in and out.

These compliances, Fig. 4.15, may be either *half-roll* or *multi-roll* and are cemented to the main body portion of the cone. They are referred to as *high compliance type (HI-C*) and, because of their looseness, permit the cone to move over abnormally large excursions.

Sometimes the *rim compliance* is accomplished by providing an annular ring of soft chamois leather, which is cemented both to the basket edge and to the paper diaphragm. A general term used for this edge compliance is the *surround* because it literally surrounds the speakers.

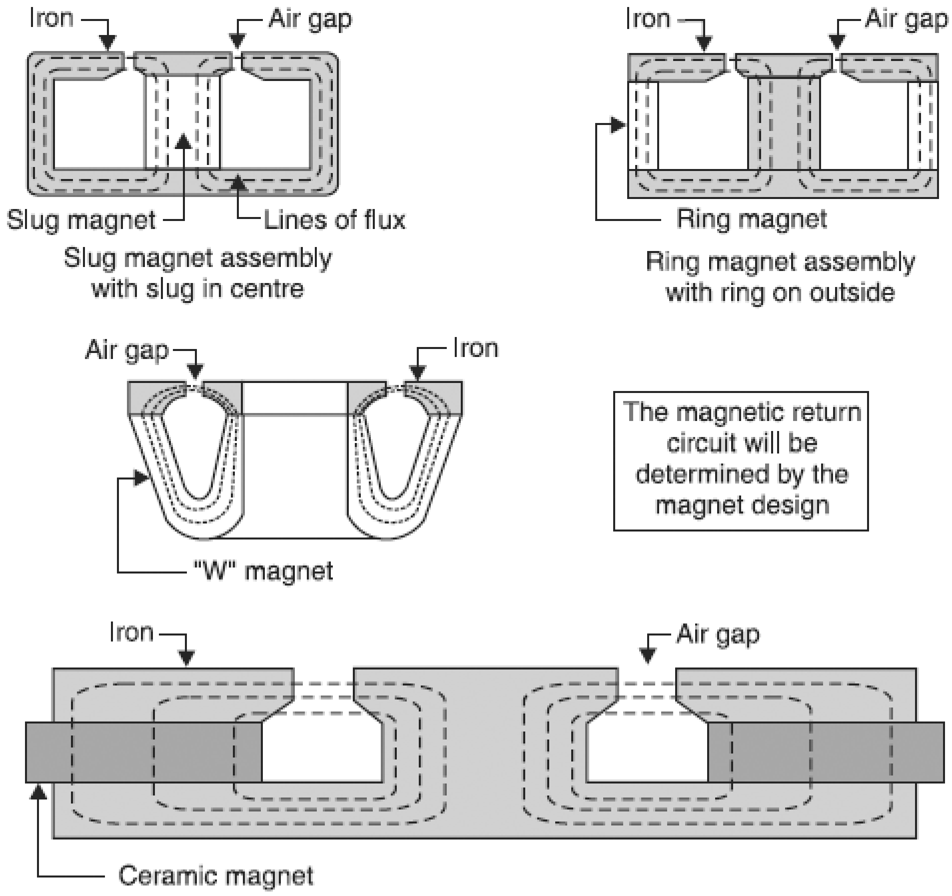


Fig. 4.14 The magnetic circuit

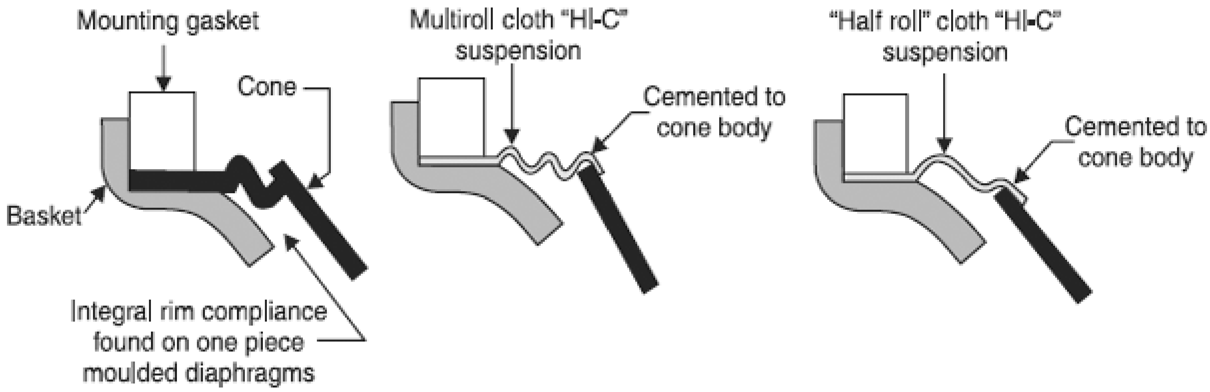


Fig. 4.15 Various types of diaphragm suspensions

LOUDSPEAKER IMPEDANCE

The impedance looking into the voice coil is not only the *self-impedance of the coil* itself but a combination of the self-impedance and the more important *reflected acoustic impedance.*

A parallel may be drawn with a transformer or motor. Each of these devices draws a small current when operating *unloaded*, indicating a relatively high input impedance. When the transformer secondary is *loaded* by an electrical resistance, or the motor shaft is coupled to a mechanical load, the input current rises and *the input impedance of each device is lowered in proportion.* In other words, the load impedance has been *reflected* into the input circuit in each case, whether it is an electrical load on the transformer or a mechanical load on the motor.

The voice coil winding is similar to any other coil in that, it has *resistance* (due to the wire used in the winding), *inductance* (due to the turns of the windings), and a small amount of *capacitance* (distributed between the turns). The resistance and the reactance of the coil combine to form the *self-impedance of the winding* without any impedances coupled into it from its association with the other parts of the speaker.

The self-impedance of the voice coil is modified by the reflected impedance of the load on the diaphragm. Mechanical inductance is called *inertance.* The degree to which the air tends to stay at rest is a measure of its inertance. It is the inductance of an electrical circuit which provides electrical inertia and it is the current which tends to stay at rest or in motion in proportion to the amount of inductance present. Mechanical inductance when applied to the air is also referred to as *acoustic inductance.* The term inertance is especially applied to the acoustic system and the air in contact with the diaphragm. The mechanical inductance of the cone and voice-coil structure and of its suspensions is also a factor in the input impedance to the voice coil, and is reflected back to it with the acoustic inertance.

Mechanical capacitance is called *compliance.* When force is applied to the spring, energy is stored in it. When the spring is released, the stored energy is also released. This is exactly what happens electrically in a capacitor in which energy is stored by the flow of current into the capacitor by application of a voltage. The applied voltage is analogous to the applied mechanical force, and the resulting current is analogous to the motion or change or displacement of the spring. In mechanical systems, we call this effect mechanical compliance. The cone suspensions act as springs and offer resistance to cone motion which increases as cone displacement increases. The suspension compliance is the main capacitive effect, although the springiness of the air load and the cone and voice coil structures during flexing add other capacitive factors. When applied to the air, this effect is called *acoustic compliance.*

Mechanical *resistance* is friction; the resistance force developed when two or more surfaces, layers, or group of particles rub together. In a speaker of the dynamic type, there are no material surfaces which rub together. Purely mechanical resistance arises in the friction within the cone and suspension materials when they flex during operation. *Acoustic load* (useful resistance component) is developed by the friction of the particles and layers of the air surrounding the cone when they bear upon each other or along the mechanical surfaces of the speaker assembly *when motion is imparted to the air* in the form of acoustic vibrations.

Mechanical components of *impedance* have the same relationships among themselves as exist among their counterparts in the electrical circuit. *Power is dissipated only by the resistive component. The inertance and compliance produce mechanical reactance which varies in the same way as electrical reactance varies.*

ACOUSTIC IMPEDANCE AND RESONANCE

The resistances and reactances of the system (including acoustic, mechanical and electrical effects) combine in the *effective impedance* looking back into the voice coil. This combination is best visualised by an *equivalent circuit* illustrated in Fig. 4.16. The *efficiency of power* *transfer* is dependent on the proportion of the impedance represented by R_A, which represents actual acoustic power dissipated in overcoming air friction and in radiating the acoustic power.

As in a purely electrical system, the capacitive reactance C_C resonates with the combined inductive effects L_V, M_C and M_A at some frequency called the *resonant frequency of the speaker.* At the resonant frequency, all reactance is cancelled out of the system and output and efficiency are far greater than what they are for other frequencies.

The curves of Fig. 4.17 show the effect of speaker resonance on response. The response *rises* to a peak at the resonant frequency, then it *falls* rapidly at lower frequencies. The shape of the response curve shows how the resonant frequency can be used as an indication of the *limit of low-frequency response.*

The size of the cone (Fig. 4.18) is important because it influences both the low-frequency response and power-handling capacity. *The larger the cone diameter is, the greater the power capacity for all low-frequency components combined and the better the low-frequency response.* However, such improvements are not necessarily derived from larger cones unless the voice coil is appropriate. The acoustic impedance offered to the cone rises as the cone is made larger; the voice coil impedance must then also be made larger for proper energy transfer and efficiency. The major portion of the audio frequency signal power is in the low-frequency components, so that the *overall power-handling capacity* is also improved.

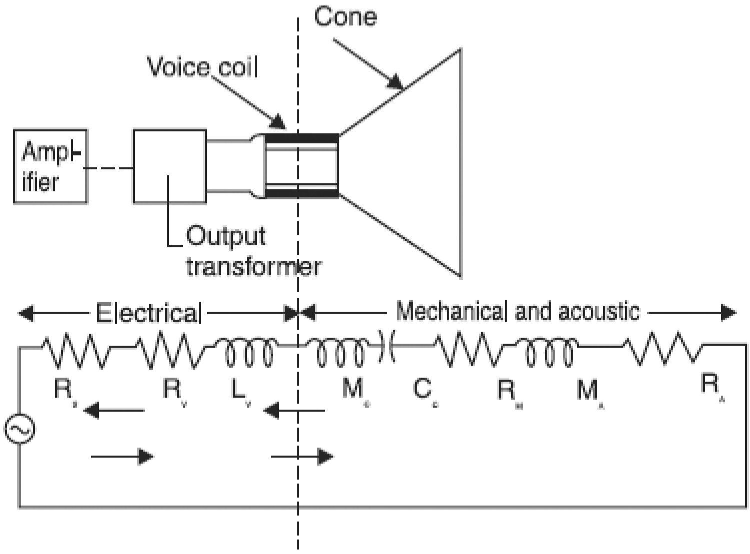


Fig. 4.16 Equivalent circuit of a loudspeaker

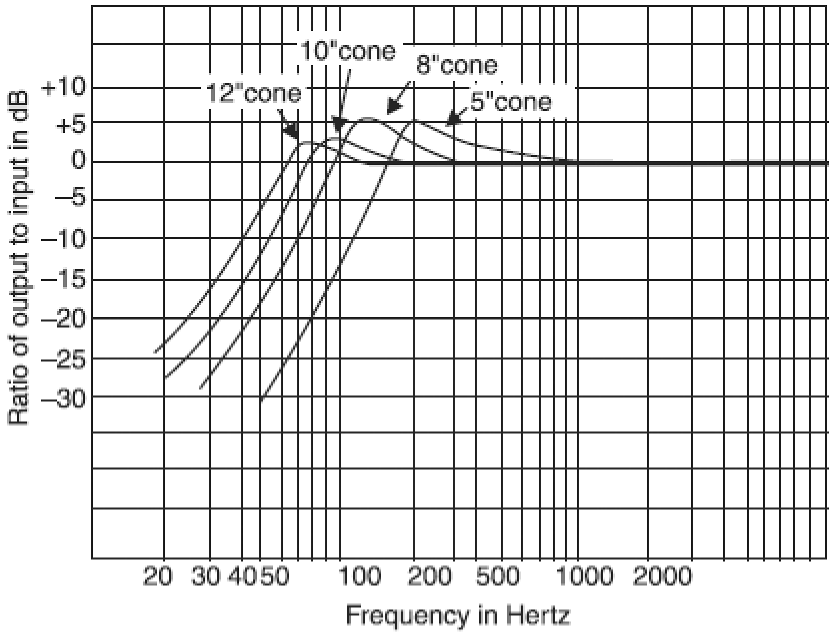


Fig. 4.17 Effect of cone size and resonance on low-frequency response

Undesired sharpness of the resonant peak can be lessened by electrical, mechanical, or acoustic *damping.* Damping is the addition of a resistive load. One method of providing damping is through proper design of the *output stage* of the amplifier. Another method of providing speaker damping is through proper design of *speaker enclosure.*

WOOFERS

There are two types of low-frequency speaker, the commonly known *woofer*, and the more recent addition the *subwoofer.* The latter is used for the reproduction of frequencies *below* those produced by the woofer, and it is generally purchased as an *add-on* to an existing system.

The low-frequency speaker provides the *bass* of any hi-fi system. Its sole purpose is to reproduce the low-frequency notes of the program source. The prime requisite for low-frequency reproduction is a *large diaphragm*, the larger the better. *The smallest diaphragm for any halfway decent woofer is 8 inches; for a subswoofer it is 12 inches.* In addition to large size, the diaphragm must be of *fairly heavy construction.* Light diaphragms just can't hold up under the vibrations encountered under the lower audio ranges.

A woofer must be able to vibrate back and forth very easily, *i.e.* have *high compliance.* One way to accomplish this is to have the diaphragm loosely connected to the frame. The gasketing that holds the periphery of diaphragm to the frame/basket is fastened so that it barely keeps the diaphragm from slipping loose, but no more as shown in Fig. 4.18. With this construction it takes less force (acoustical power in this case) to move the diaphragm any particular distance (less power required for a given excursion).

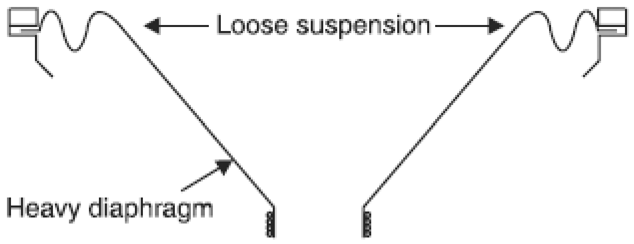


Fig. 4.18 (a) Suspension of woofer diaphragm. (b) For good reproduction a woofer should have a large diameter, stiff diaphragm, high compliance and a large voice coil.

Rather than the loose suspension system, the cone is supported by a very flexible material so that it can be moved very easily by the voice coil. *The suspension is tight but the sine wave at the diaphragm edge is made very flexible.*

The large speakers have more extended lows, the smaller ones more extended highs.

A woofer must also have a *large voice coil* to handle considerable heat. The larger the voice coil, the more the current produced by the amplifier output circuit and, therefore, the more the power the woofer can handle.

Finally, a *strong magnet* can be of great help to move the heavy voice coil and cone assembly too well. The better the woofer, the heavier the magnet assembly (unless it's ceramic).

To sum up, *a good woofer must have a large, heavy diaphragm, a strong magnet, high compliance, and a large voice coil.*

MID-RANGE AND EXTENDED-RANGE SPEAKERS

The *mid-frequency loudspeaker* is supposed to have a good response at the frequencies located between those covered by the woofer (low-frequency speaker) and the tweeter (high-frequency speaker). However, some *extended-range speakers* can be used as general purpose units. An adequate **general purpose speaker**, Fig. 4.20 must have good response at both the low and the high ends of the audio range, not to mention in the middle.

Tests by various manufacturers reveal the following data: A *15 inch* diameter loudspeaker will give excellent low-frequency response but will be sadly deficient in the highs. A *12 inch* diameter unit will still give good low-frequency response, halfway decent middle frequency response, but not too good treble response. *The 8 inch unit, one of the most popular today, is the smallest that will permit fairly good response at low frequencies, and good response at the middle and high frequencies.* This is shown in Fig. 4.19.

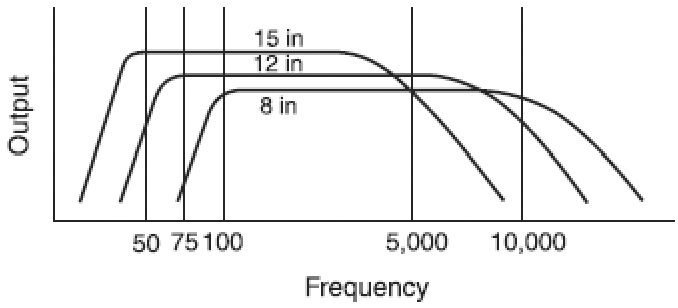


Fig. 4.19 The popular sizes of speakers have overlapping characteristics of response.

Conventional extended range speakers are now available in frequency range 45 to 15,000 Hz, although *they need a good enclosure to cover the range.* To obtain this type of response, the loudspeaker must be designed in away that the diaphragm will vibrate well at the low and middle frequencies, yet have some sort of *compensation* to permit good response at high frequencies. A three-way speaker system is shown in Fig. 4.21.

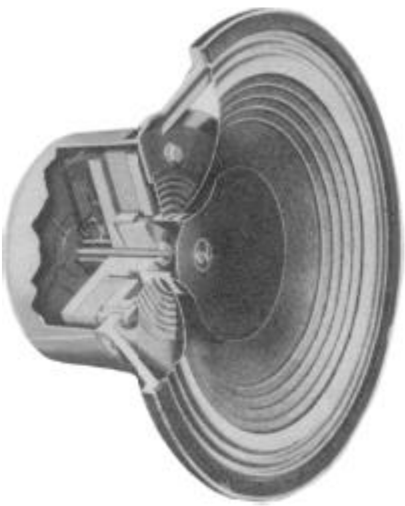


Fig. 4.20 Cut-away view of a general-purpose speaker

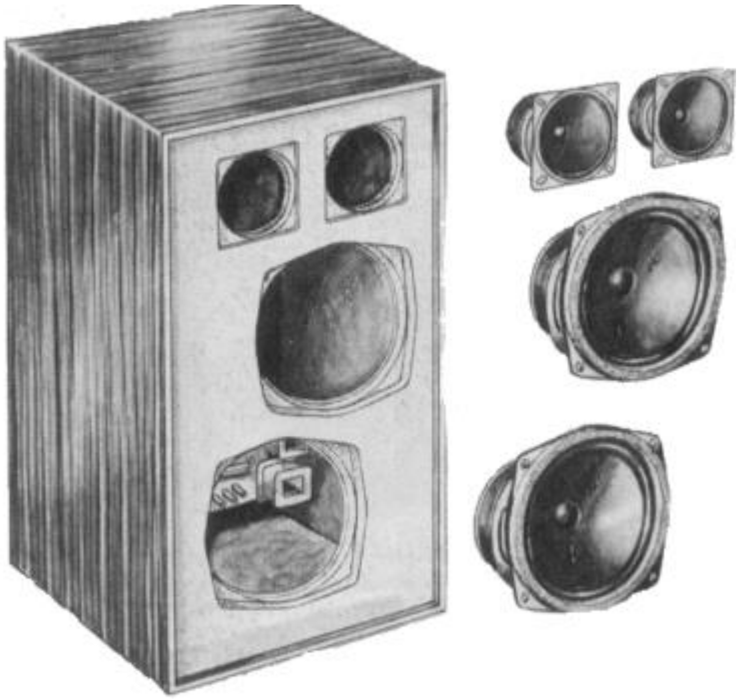


Fig. 4.21 A three-way speaker system

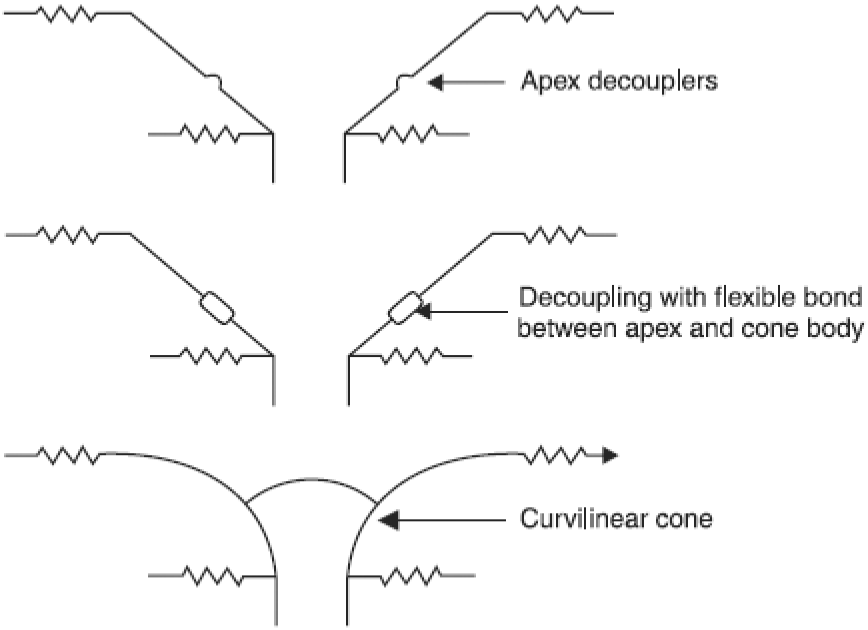


Fig. 4.22 Various diaphragm apex decoupling methods. By using mechanical dividing methods, the apex of the cone can be separated to handle high frequencies while the cone vibrates as a single unit for the lower ones.

The top illustration in Fig. 4.22 shows the converse type of diaphragm. Here the curvature of the diaphragm allows the apex to *isolate* itself without the need for special *decoupling devices.* The centre diagram shows the *apex decoupler* built right into the apex area. The lower diagram illustrates the joining of two discrete diaphragms to each other by means of a *flexible connection.* Thus one diaphragm can vibrate while the other one doesn't.

In the *duplex* type or *coaxial speaker*, we have a high-frequency horn centered within the framework of the low-frequency speaker, as shown in Fig. 4.23. However, *there is only one voice coil;* thus it more properly falls into a **single-unit extended-range** category. The switch from high to low-frequency response is made by a *mechanical crossover* built right into the speaker framework.

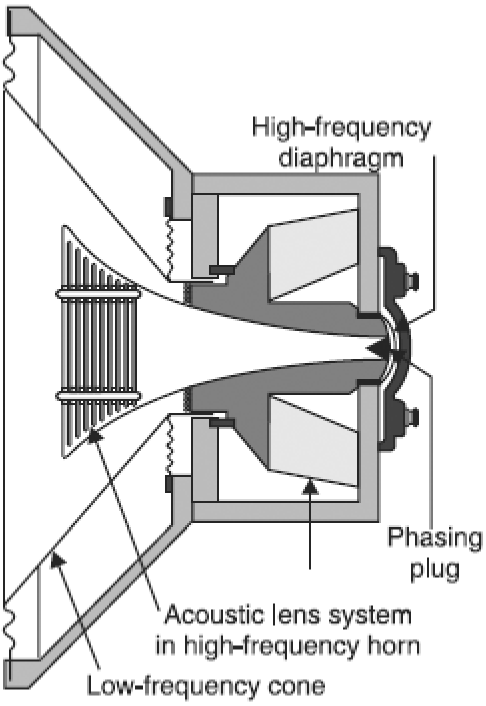


Fig. 4.23 A dual-concentric speaker

HIGH FREQUENCY LOUDSPEAKERS

There are two main types of high frequency speakers; the well-known *tweeter* and the more recent *supertweeter.* Supertweeters can be add-ons or they can be *integral with the system.* Six basic high-frequency speakers (tweeters) exist.

1. The *cone* is a physically disincentive version of the woofer.
2. The *dome,* so called because of its dome-shaped diaphragm.
3. The *horn,* so named because it is a horn.
4. The *Heil air-motion transformer* which uses the principle of lever in its operation, named after its inventor, Dr. Oskar Heil.
5. *High polymer molecular-film* tweeter, uses the piezoelectric effect for its principle of operation (used exclusively by Pioneer).

6. The *electrostatic* tweeter works on the principle of attraction or repulsion between two metal plates.

CONE TYPE TWEETERS

Since *tweeters* must reproduce high-frequency notes, they must resonate at high frequencies. *High resonant frequencies are obtained with lightweight, stiffly supported mechanisms.* To make the diaphragm of a cone-type tweeter light, it must be small. When we reduce the size and weight of the diaphragm we must, in turn, reduce the size of the *voice coil* also. Luckily, high frequencies carry only a comparatively small amount of electric power, therefore, *the small voice coil is not subjected to electrical overload.* Without exception, it is wound with light weight wire such as aluminium wire or ribbon. The lightness of the moving system provided by aluminium makes the high frequency response much better than if copper were used.

Cone type radiators tend to concentrate radiation of the high frequency components of sound in a narrow cone about the axis of the radiator. The degree of directivity of speaker is indicated by a directivity pattern in Fig. 4.24. The axis of the radiator is considered the *reference line* with an angle of zero degrees. *Directivity patterns* are normally shown as a top view in the horizontal plane through the radiator axis. A cone in free space should have the same pattern in a vertical plane.

The line *OA* in Fig. 4.24 indicates by its length that the sound radiated along it is a *maximum* in comparison to that in any other direction. At an angle 45°, the line *OB* is a measure of the relative sound intensity in that direction. Since *OB* is only half as long as *OA*, a listener along *OB* would listen only about half the volume compared to what a person along *OA.* At angles near 90°, the pattern indicates *minimum* (zero) radiation. In any practical setup, such a *zero area* would not exist because sound waves reach there by *reflection.*

Because, *directivity normally varies considerably with frequency*, a complete diagram (Fig. 4.25) must show separate patterns for each of at least several frequencies. Fig. 4.25 depicts variation of directivity with frequency for a 12-inch cone, assuming that the speaker is mounted in an infinite baffle. *Notice how much narrower the radiation pattern is at highs than at lows.*

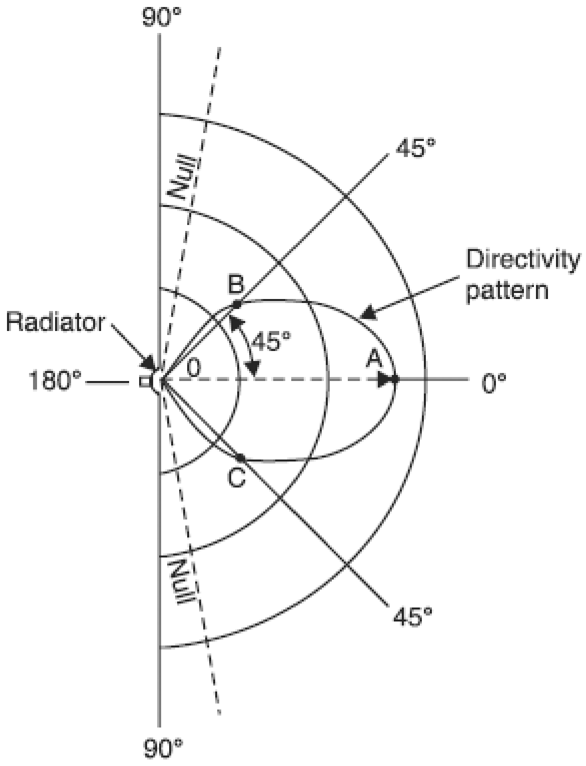


Fig. 4.24 Radiation pattern for a typical cone at one frequency

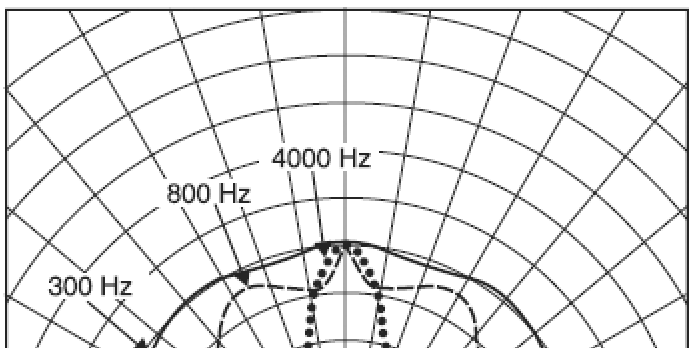


Fig. 4.25 Variation of directivity with frequency for a typical cone radiator

A single cone-type tweeter distributes high-frequency sounds unevenly. It *lobes* the higher frequencies directly out in front and tends to cause a drop off at the sides. This effect can be overcome by arranging two or more cone tweeters as shown in Fig. 4.26. In this way, *overlapping individual lobes from separate speakers cover the listening area.*

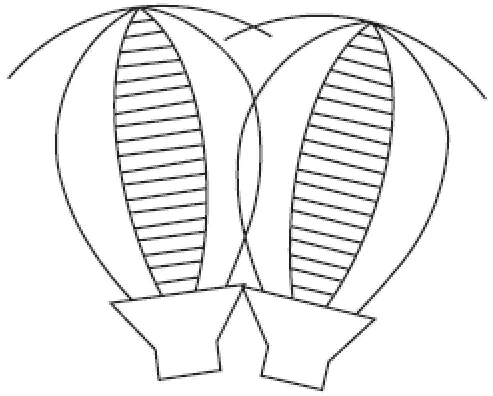


Fig. 4.26 Sound distribution of two-tweeter array. Highs can be spread out by angling the speakers

DOME TYPE TWEETERS

Uniformly dispersed flat energy response begins with a speaker system's ability to radiate sound at all frequencies evenly in all directions. *Even dispersion* of sound energy means that the sound emanating from the program source will be heard same by listeners in all parts of the room.

For *low frequency sounds* this problem of dispersion is not of practical consequence, since they are very nearly omnidirectional. The limiting factor for high-frequency sounds is that a speaker will begin to be directional when its circumference equals the wavelength of the frequency being reproduced. *Directionality increases as the wavelength decreases with respect to the speaker's dimensions.* The laws of physics dictate the most direct approach to the problem of even dispersion of high-frequency energy; *the drivers used must be as small as possible.* Dome tweeters, Fig. 4.27, are designed according to this principle in order to use these physical laws to the listener's advantage.

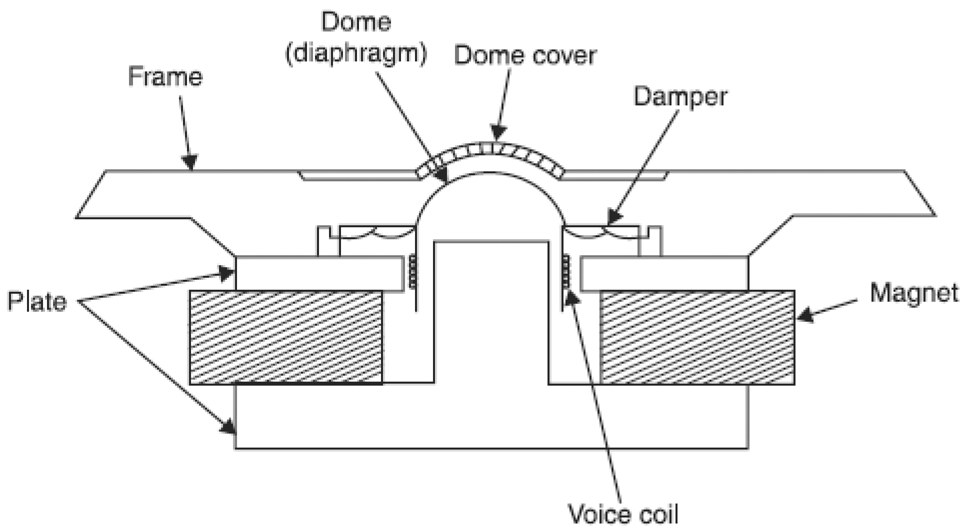


Fig. 4.27 Dome-type tweeter

HORN TYPE TWEETERS

To obtain reasonable output from a loudspeaker, we must vibrate large amounts of air. For this, we must usually have fairly large vibrating surfaces, such as the cones in woofers. The larger the cone surface, the greater the output. But the tweeter's cone (diaphragm) must be small to attain its high-frequency response. Thus only a small amount of air can be moved, reducing the output power.

We can increase the acoustic output from any type of diaphragm if we couple directly to a horn, converting the system to a horn-loaded one. Fig. 4.28 shows the relative difference in size between the diaphragms of a cone-type tweeter and a horn-loaded one. The driving force of the voice coil of the latter is distributed between the small mass of the diaphragm and the mass of air in the horn. Since air weighs much less than paper or metal the overall load on the voice coil, for the same acoustic output as that of the cone type tweeter, can be greatly reduced. Also, *for the same electrical input, the output of the horn loaded system is greater*.

A *horn* is a tube so flared (tapered) that the diameter increases from a small value at one end called the *throat* to a large value at the other end called the *mouth*. Horns, Fig. 4.28 (b), have been used for centuries for increasing the radiation of the human voice and musical instruments.

The horn does acoustically what the cone does mechanically. It couples the small voice coil area to a large area of air. In this way, the horn acts as an *acoustic transformer* and converts the relatively high impedance at the throat and driver.

The horn is a fixed physical boundary for its enclosed column of air and does not vibrate itself. Acoustic energy fed to its throat must therefore be obtained from a vibrating diaphragm which converts mechanical motion from the driver voice coil to acoustic energy. Although *the cone-type radiator acts as both diaphragm and radiator and transducer from mechanical to acoustic energy, the horn acts only as a radiator, with both input and output energy being acoustic.*

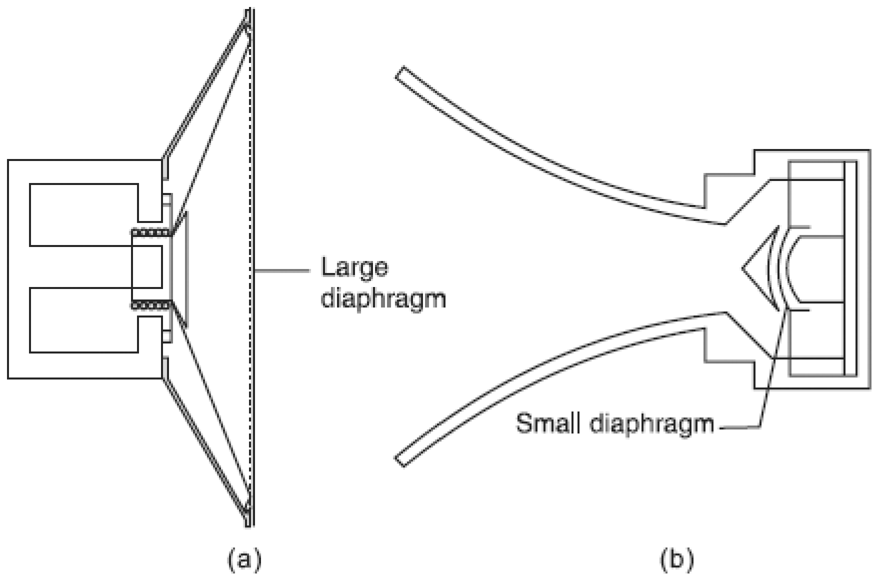


Fig. 4.28 Cone and horn-loaded tweeters (a), when used as a direct radiator will produce approximately the same sound output as the horn radiator (b) with approximately only 20 percent of the direct radiator power because of the small light diaphragm of the latter's compression driver.

For maximum efficiency, horn driver units, such as the one shown in Fig. 4.29, are of the *compression type.* The back of the driver is completely sealed to provide a stiff air cushion behind the diaphragm. The *sealed-in stiff air cushion* is the underlying principle of acoustic suspension, or air-suspension speaker systems.

Any device that radiates energy into three-dimensional space has certain very specific *directional properties.* For example, an unshaded light bulb is pretty much omnidirectional, casting illumination in most directions. Like lamps, speakers have definite directional characteristics. At very low frequencies, any speaker is virtually omnidirectional. *The directional characteristics of a speaker change with the frequency of the sound being generated, particularly with how the wavelength of the sound relates to the physical size of the speaker's diaphragm.* For a flat, circular cone, dispersion will be virtually omnidirectional for frequencies with wavelengths that are more than 4 times the diameter of the cone. Dispersion narrows to approximately 60° when the wavelength equals the diameter, and to 30° when the wavelength is half the diameter. A speaker with a 2 - inch radiating surface will be practically omnidirectional at 1,000 Hz, but will radiate a 10,000 Hz tone in a beam not much more than 30° wide.

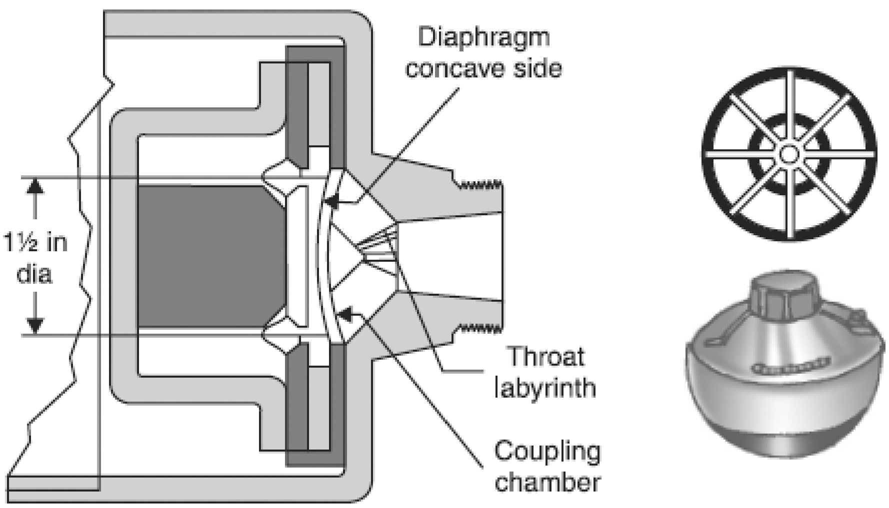


Fig. 4.29 Cross-section of a compression horn driver unit

Horns can be, and are, designed for controlled wide-angle distribution. Such horns are square or rectangular. These horns achieve similar results in different manners.

The *diffraction horn*, Fig. 4.30, operates on the principle that sound coming from a narrow-slit, which is small compared to the wavelength, emerges in a cylindrical wavefront pattern from the slit. That is, *the wavefront diffracts out of the narrow slit.* This type of horn has smooth angular response with no lobes or valleys in the sound pattern.

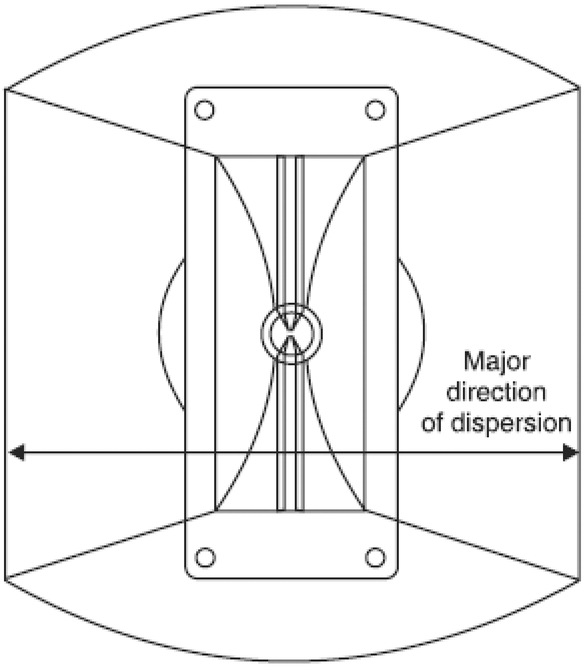


Fig. 4.30 The mouth of a diffraction horn is small compared with wavelength, making it a horizontal-dispersion slit source.

The *reciprocating flare horn*, Fig. 4.31, distributes the sound over a wide angle by *reversing* the direction in which pressure builds up within the horn. Its mouth construction is designed for horizontal dispersion. The horn first expands rapidly in the vertical direction, with practically no horizontal expansion. The sound pressure travelling along such a channel finds it relatively easy to expand vertically. But in trying to expand horizontally, the sound pressure builds up along the side walls of the horn. The result is that *the sound comes out of the horizontally shaped mouth (i.e., the reciprocating flare), with exceptionally wide horizontal dispersion.*

The *sectoral/multicellular horn*, Fig. 4.32 is similar to the reciprocating flare horn except that the mouth of the horn is compartmentalised to diffuse the sound over a wider area. What we have here is a sort of *shower head* that spreads the sound much like a shower head which spreads the water coming from it.

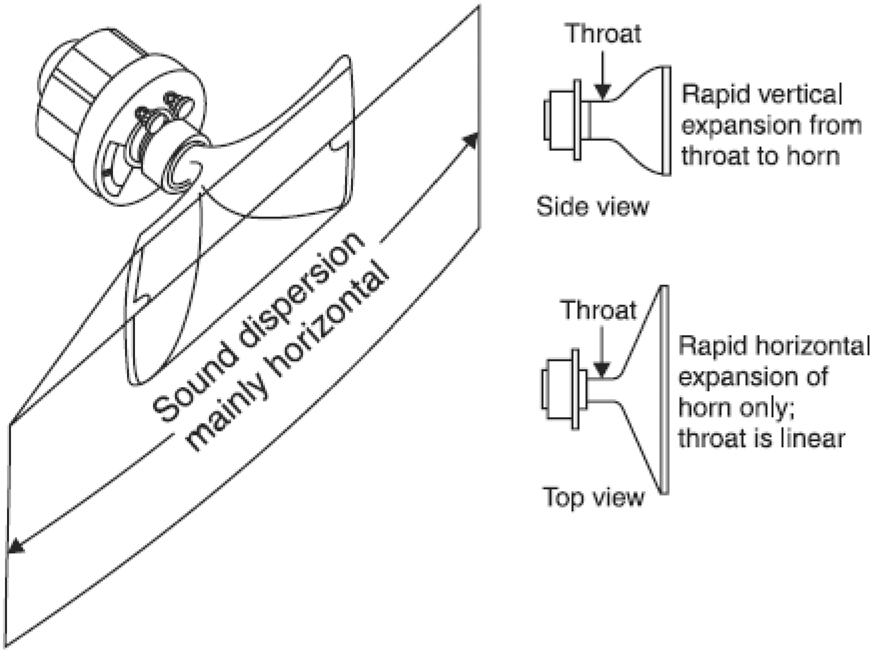


Fig. 4.31 The reciprocating flare horn looks like a diffraction type but distributes sound in an opposite manner.

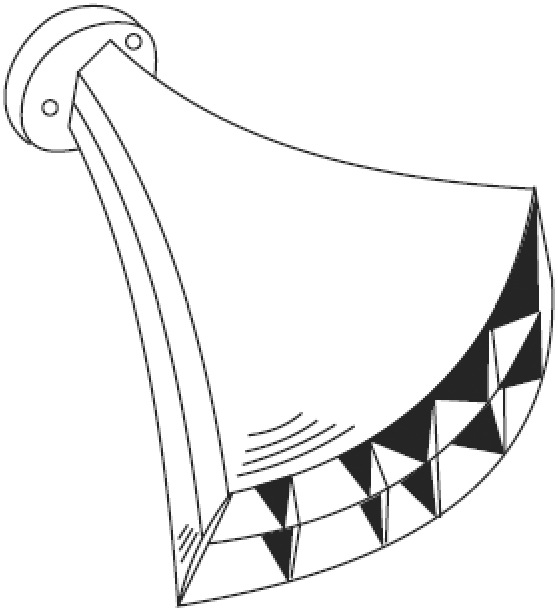


Fig. 4.32 The multicellular horn

HIGH FIDELITY

There are several schools of thought concerning high fidelity - *'play it exactly as performed'* school, *'play it so it sounds real'* school, and *'play it the way I like to hear it'* school. There is logic in all three approaches. We must accept that there are as many types of high fidelity as there are listeners. In order to satisfy these concepts, it is necessary to provide a sufficient variety of component parts to make possible many different kinds of system. We have discussed speakers from the point of view of their general application to the hi-fi field.

There is no exact scientific operational definition of a hi-fi system as yet. Standards and specified measurements of performance of a system have not been possible to establish because of the *limitations* of the human ear and because of *variations* in human taste, room acoustics, system distortions, noise and comparative volume levels.

A commonly accepted concept of hi-fi-sound is that it is reproduced sound with a high degree of similarity to that of original or live sound that has travelled from a source and has undergone several conversions through a system or several systems. Hi-fi is felt to be achieved when the sound that is reproduced has negligible distortion from the original, when it has little extraneous noise, and when the volume levels and acoustical effects are pleasing to hear.

Reproduction of sound is something like a photograph. The picture cannot carry the original scene to the viewer in every detail. Some feature of the picture may be *de-emphasised* whereas other features may be *emphasised* intentionally, or distortion may be introduced for *purely aesthetic reasons.* Distortion of this type can greatly improve the illusion that the photographer is trying to create. In the same way, the picture can be spoilt by undersirable distortions and effects, such as poor focus, poor film and improper lighting.

Like photography, modern hi-fi techniques encompass *controls for modification of the original (live) sound to compensate for certain defects and make provision to actually improve the effects according to an individual-listener's tastes.* Undersirable distortions, differences in comparative sound levels, and injection of extraneous noise are also held to a minimum so that the pleasing qualities of the original sound will not be reduced.

In addition, modern concepts of hi-fi take into consideration the listener, his ear mechanism, and his nervous response, plus his listening experience and training.

The word presence is used to describe the degree of realism of the reproduced sound. This term suggests that the reproduction is so real that the listener can feel the presence of the source that is causing live sound, even though the source may be extinct.

A complete *hi-fi system* may be divided into functional sections as illustrated in Fig. 4.33. The way in which the output differs from an input or a desired ideal output is called *distortion.* Distortion may be created in anyone or more of these sections. If more than one section is causing distortion, the final output sound may reflect the *sum of the distortions from all distorting sections.* A section may be purposely designed to *introduce* distortion of a type which *compensates* for inherent distortion in another section. For example, bass and treble boost circuits can be used to offset (at least partially) the falling off of the response of a speaker at the highest and lowest frequency portions of the response range.

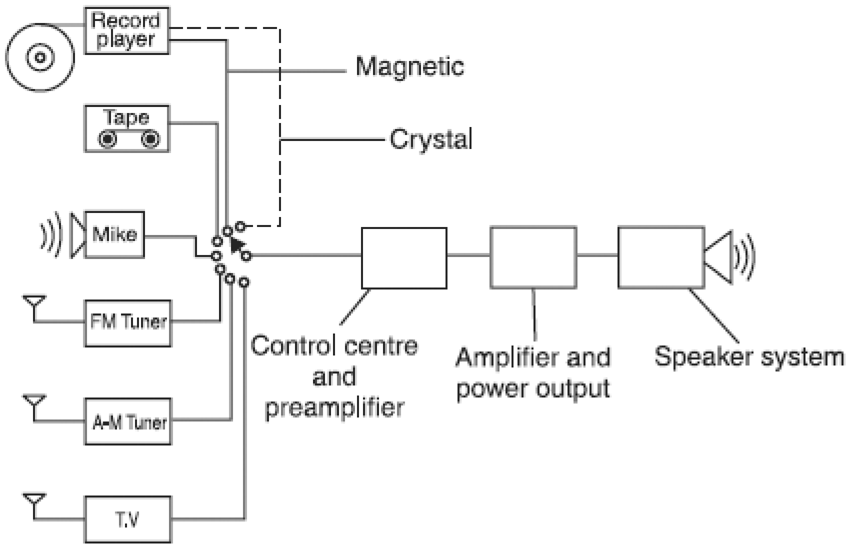


Fig. 4.33 Block diagram of a high-fidelity system

The hi-fi system is somewhat like a chain, which is likely to be limited in overall performance by its weakest section; but the chain analogy breaks down in the foregoing case in which *the distortion introduced by one section may be used to compensate for distortion in another section.*

The *speaker system* is the weakest link in the hi-fi equipment chain because two conversions of energy must take place, electrical-to-mechanical-to-acoustic. Such energy conversion is known as transduction, and the devices which effect it are known as *transducers.*

Input devices such as phono pickups and microphones are also transducers, and have many of the same weaknesses as speakers, though to a lesser degree, because of the relatively *low power levels* at which they operate. Input devices provide transduction between sound input (or physical motion of a phono-pickup needle) and electrical output, just the reverse of the action in speakers.

The *amplifier system* can also contribute distortion. The *voltage amplifier* stages are the least troublesome. The *power amplifier* stage is an important contributor to the overall distortion in the system.

Keeping in mind the types of imperfections and distortions they can introduce, we may now summarise the features of a *theoretically ideal system.*

1. Interpret, amplify, compensate and reproduce sound components of *any and all frequencies* in the audible range with good efficiency.
2. Add negligible frequency components *not* in the original sound.
3. Distribute the sound in such a way that its sources would appear to be located *nearly* the same as they were in the original and so that the *quality* of the sound would be independent of the *location* of the listener with respect to the speaker system.
4. Allow negligible *unnatural delay* of some frequency components relative to others.

5. Reproduce *without resonance effects* or hangover, the sudden large changes in sound volume level.

MULTISPEAKER SYSTEMS

We will now concentrate on the more specific and specialised units that become component parts in the more expanded hi-fi reproducer systems.

Multispeaker systems have much to offer for good hi-fi reproduction in the way of characteristics that are virtually impossible to obtain from a single wide range speaker *Advantages* derived from multispeaker systems are due to the fact that with two, three, or four speakers in the reproducing system, *we have better control of the overall performance characteristic of the system control of the individual component speakers.*

The situation may be likened to the difference between having only one ceiling lighting fixture in a room to provide overall illumination and having several lamps in corners and on tables, to provide more adequate light coverage in *specific areas* where light is most needed and in *degrees* best suited to those areas.

Since multispeaker systems, Figs. 4.34 and Figs. 4.35, are composed of combinations of special purpose speakers, *the multispeaker system as a whole is a very efficient system.* Multispeaker systems have the following advantages :

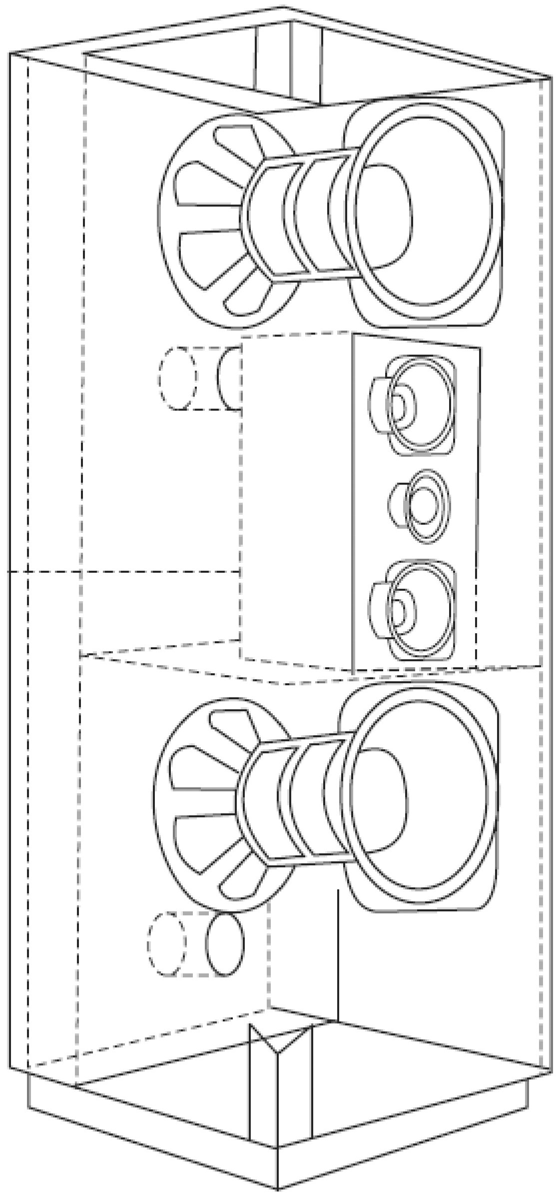


Fig. 4.34 Each cabinet of the multispeaker system is divided into five separate chambers, of which two are of the band-pass bass reflex type.

Fig. 4.35 A typical multispeaker system

1. Multispeaker systems have *reserve power handling capacity* necessary for high program bursts.
2. Multispeaker systems *reduce intermodulation distortion.*
3. Multispeaker systems *may be balanced one against the other by means of volume controls* to give that particular feeling of concert hall reality that most pleases the listener.
4. A *smooth overall response* may be obtained from multispeaker systems.
5. Multispeaker systems have compatible '*roll off' or 'cut off' characteristics;* speaker ranges overlap.
6. Multispeaker systems provide *flexibility of performance.*

CROSSOVER NETWORKS

In multispeaker systems, in which specialised speakers are used for different frequency hands, *it is necessary and desirable to segregate different bands of frequencies into the respective speakers designed to handle them.* This segregation of the various bands of acoustic energy ensures *optimum utilisation of audio power* resulting in better overall performance of the system.

The simplest type of network consists merely of a single *capacitor*, as illustrated in Fig. 4.36 (a). The fact that the reactance of a capacitor is inversely proportional to frequency is employed to distribute the audio signal. The tweeter and woofer voice coils are connected in *series* and a capacitor is connected *across* the woofer. The value of capacitance is made such that frequencies above the range of the woofer, the reactance of *C* becomes so low that it shunts the woofer, which acts as a bypass capacitor.

Low-frequency components can be kept out of the tweeter if a *parallel* connection of voice coil is used with capacitor in *series* with the tweeter circuit, as shown in Fig. 4.36 (b).

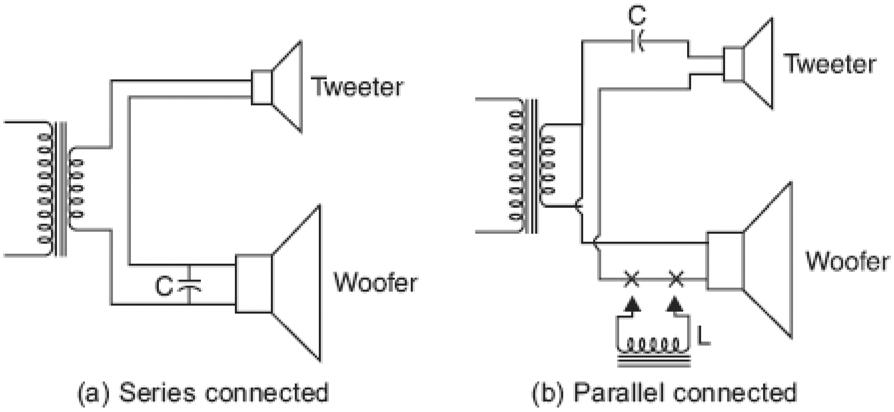
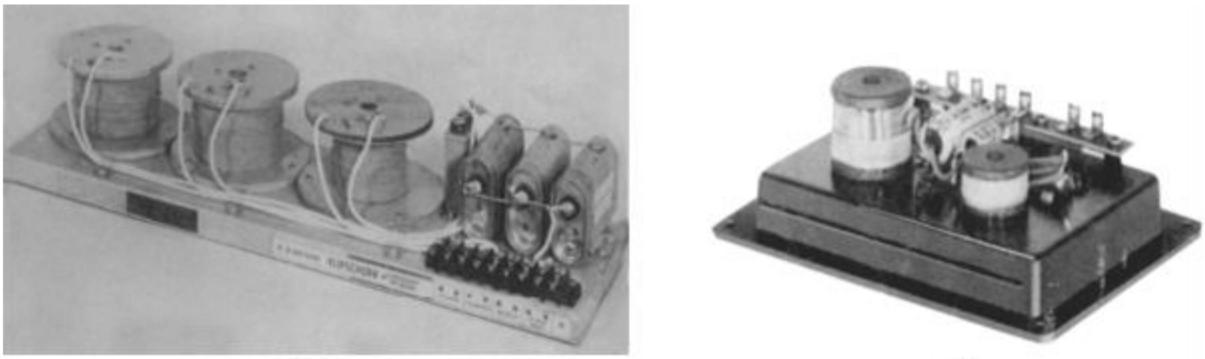


Fig. 4.36 Simple divider circuits employing single reactances

Inductances can be used along with capacitors to make the crossover network more complete. For example, in Fig. 4.36 (b) the inductance *(L)* can be connected in series with the woofer leads as shown. The inductance, the reactance of which increases with frequency, *chokes* the high frequency components out of the woofer, and the capacitor (*C*) *blocks* low frequency components out of the tweeter. The values of *C* and *L* must be such that the reactance in each case is a little lower than the voice-coil impedance in the frequency range to be attenuated.



(a) (b)

Fig. 4.37 Typical crossover networks

Although the crossover range should not be too narrow, simple reactance circuits are ordinarily too broad in the change over region. A combination of *low pass filter* (for the woofer) and a *high-pass filter* (for the tweeter) is usually employed. *With this type of circuit, much more rapid attenuation can be made near the crossover frequency than is possible with simple capacitor and inductor arrangements.*

Attenuation of 12 dB per octave is considered proper in most applications. Gradual crossover arrangements attenuate at about 6 dB per octave. Filters with sharper cut off than this can be constructed by the use of additional components, but *power losses* in the filter become excessive and the additional sharpness is not necessary anyway.

A typical crossover network response graph is shown in Fig. 4.38. The curve of woofer output crosses the curve of tweeter output response at the crossover frequency. This intersection is at the 3 dB or *half-power level*, at the crossover frequency, half the output power is being fed to each unit. The individual response characteristics of the woofer and tweeter must *overlap* substantially.

One of the most popular types of crossover networks is shown in Fig. 4.39 alongwith the formulas for calculating the values required for any crossover frequency f_c and speaker impedance Z.

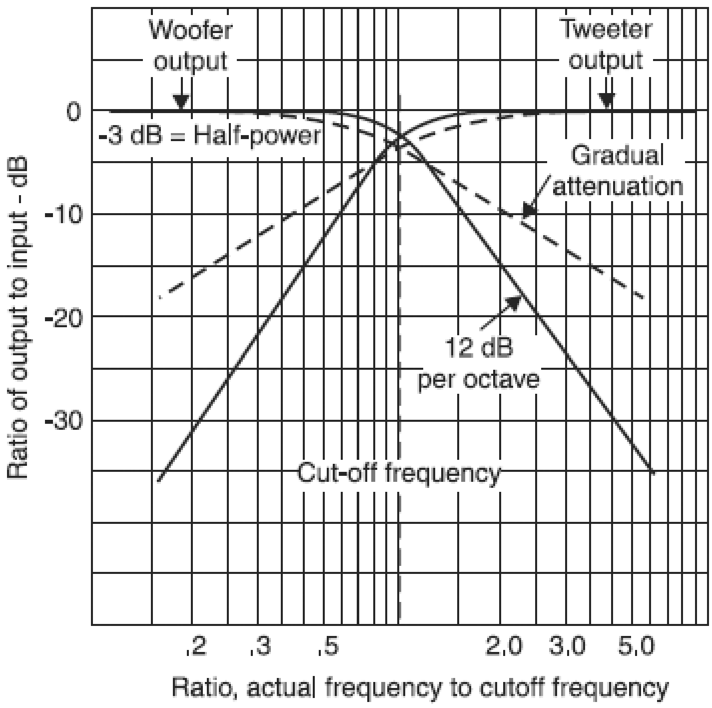


Fig. 4.38 Crossover network reponse curves

Example 4.1 Find the component values for f_c = 500 Hz and Z = 8 ohms.

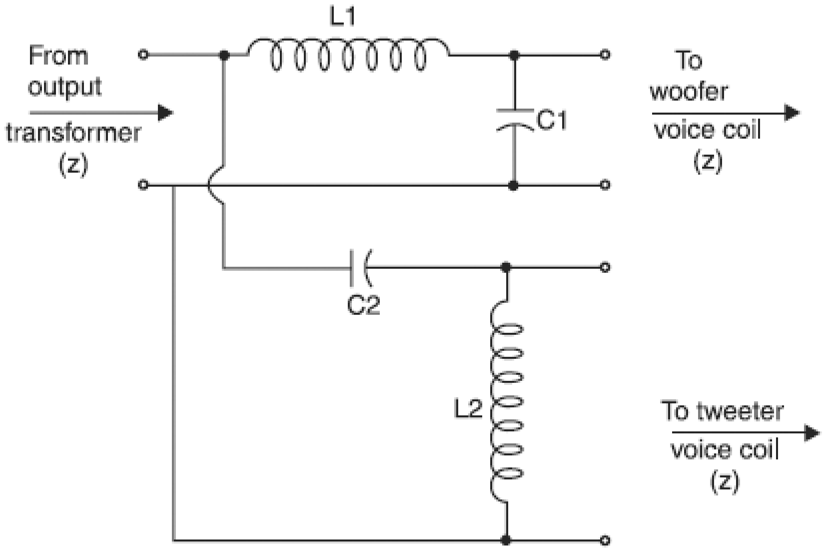
Solution

$$L_1 = \frac{0.255 \times 8}{500} = \mathbf{4.08\ mH}$$

$$L_2 = \frac{0.15 \times 8}{500} = \mathbf{2.4\ mH}$$

$$C_1 = \frac{0.15}{500 \times 8} = \mathbf{37.5\ \mu Fd}$$

$$C_2 = \frac{0.10}{500 \times 8} = \mathbf{25.0\ \mu Fd}$$



$$L1 = \frac{0.255Z}{f_c} \text{ Henrys} \qquad C1 = \frac{0.15}{f_c Z} \text{ Farads}$$

$$L2 = \frac{0.15Z}{f_c} \text{ Henrys} \qquad C2 = \frac{0.10}{f_c Z} \text{ Farads}$$

Fig. 4.39 A filter type crossover network

IMPEDANCE MATCHING

It will be useful to give a simple recapitulation of transformer theory. A transformer represents a straight forward application of the phenomenon known as *electromagnetic induction*, whereby any change in the magnetic flux linking an electric circuit is accompanied by an induced electromotive force in that circuit. The *value* of this induced voltage will, at any instant, be proportional to the rate of change of the flux, while its *direction* will always be such as to oppose the change producing it. The change of linkages may be produced by movement of the electric circuit with respect to a magnetic field which is constant, as in a dynamo, or by varying the current in a stationary electric circuit. When the induced emf produced in one circuit is due to a current varying in an adjacent circuit, as in a transformer, it is said to be due to the *mutual inductance* between the two circuits. The emf induced in a circuit by a varying current flowing in the circuit itself is said to be due to the *self-inductance* of the circuit. The inductance of a conductor can be *increased* by winding it in a coil and *further increased* by winding the coil on a core of ferromagnetic material.

In general, a transformer consists of two insulated coils of wire so associated that the magnetic flux due to a current in one coil is effectively linked with the turns of the other coil. The coils are normally wound on a core of ferromagnetic material which greatly increases the magnetic effect of the current. Usually one of the windings, the one to which the input is connected, is called the *primary* and the other is called the *secondary.* In some instances additional windings may be provided to give alternative outputs.

An elementary transformer corresponding to the form used by Faraday in his original experiments is shown diagrammatically in Fig. 4.40. An alternating current, *I* flowing in the primary winding will cause an alternating magnetic flux, indicated by the dotted lines, to circulate in the ring core. In doing so, this alternating magnetic flux will link with the turns of both windings and induce, assuming an ideal transformer, an emf, E, in each turn, proportional from instant to instant to the *rate of change* of current in the primary. This induced voltage will act in such a direction so as to tend to *oppose* the change of flux and in the primary winding it will take the form of a *back emf*. If N_p is the number of turns in the primary winding the total back emf will be equal to N_pE and neglecting any losses in the transformer this will be equal to the impressed emf E_p In other words, stability is established when the primary current has a value equal to that required to induce a back emf, opposite to the applied emf.

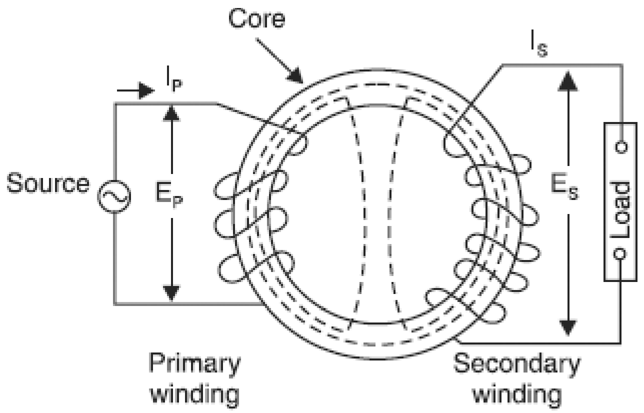


Fig. 4.40 Elementary transformer

Similarly, if N_s is the number of turns in the secondary winding, the total emf, E_s, induced in the secondary winding is equal to N_sE. We thus have

$$E_p = N_p E \qquad ...(4.1)$$

and

$$Es = N_s E \qquad ...(4.2)$$

and therefore

$$\frac{E_p}{E_s} = \frac{N_p}{N_s} \qquad ...(4.3)$$

Again assuming an ideal transformer having no losses, the power in the secondary circuit will be the same as that at the input to the primary winding; thus we have.

$$I_p E_p = I_s E_s \qquad ...(4.4)$$

so that

$$\frac{E_p}{E_s} = \frac{I_s}{I_p} \qquad ...(4.5)$$

We therefore have the following relationships :

$$\frac{N_p}{N_s} = \frac{E_p}{E_s} = \frac{I_s}{I_p} \qquad \ldots(4.6)$$

Thus the ratio of the primary and secondary voltages is equal to the turns ratio, while the ratio of the primary and secondary currents is the inverse of the turns ratio. The transformer, therefore, provides a ready means of converting an alternating supply voltage to another value, *either higher or lower*, depending on the requirements, and this constitutes one of its most common uses. Where ac mains are used to provide power supplies to telecommunication equipment, transformers also serve the purpose of *isolating* the equipment from direct connection to the mains, both on grounds of safety and to prevent unwanted earth potentials reaching the equipment.

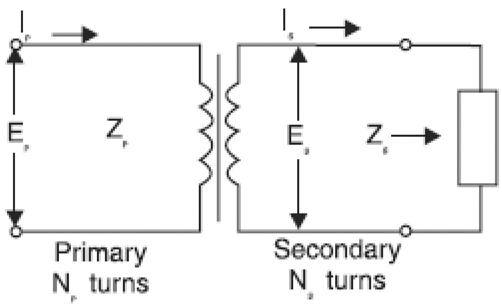


Fig. 4.41 Impedance transformation

Apart from their uses in connection with power supplies, transformers find very many uses in telecommunication circuits. In signalling and switching circuits they may be used for isolating two parts of a connection for direct current while permitting the transmission of speech currents between the two parts. They may also be used to connect an earthed circuit to one which has both sides *balanced* with respect to earth.

A further use of a transformer is for *impedance matching.* This is to allow a low impedance circuit to be connected to a high impedance circuit or vice versa without loss of power. In this case *the turns ratio required is equal to the square root of the ratio of the impedances to be matched* as can be seen from the following:

Let Z_p = Impedance as measured at terminals of primary winding (Fig. 4.41)

Z_s = Impedance of circuit connected to secondary winding

Then $\frac{E_s}{I_s} = Z_s$ and $\frac{E_p}{I_p} = Z_p$

Therefore
$$\frac{Z_s}{Z_p} = \frac{E_s}{I_s} \times \frac{I_p}{E_p} = \frac{E_s}{E_p} \times \frac{I_p}{I_s} \qquad ...(4.7)$$

But
$$\frac{E_s}{E_p} = \frac{I_p}{I_s} = \frac{N_s}{N_p}$$

Therefore
$$\frac{Z_s}{Z_p} = \left(\frac{N_s}{N_p}\right)^2 \qquad ...(4.8)$$

So that
$$\frac{N_s}{N_p} = \sqrt{\frac{Z_s}{Z_p}} \qquad ...(4.9)$$

Descriptive Questions

1. Explain the difference between a microphone and a loudspeaker.
2. What are the requirements of an ideal loudspeaker?
3. What are the different types of loudspeaker? Explain the working of any one of them with the help of a neat diagram.
4. Discuss the limitations of ceramic and electrostatic speakers. What are their fields of application?
5. Draw the equivalent circuit of a dynamic loudspeaker and explain the significance of each component.
6. Explain the following terms :
    a. Compliance
    b. Inertance
    c. Reflected impedance
7. What type of voice coil and voice coil suspension will you use in
    a. Woofers
    b. Tweeters
8. Explain in detail the working of a permanent magnet loudspeaker.
9. What is the difference between a cone-type tweeter and a dome-type tweeter?
10. Explain why the efficiency of an indirect radiating loudspeaker is greater than that of a direct radiating loudspeaker?
11. Discuss the relative merits and demerits of horn type and cone type speakers.
12. Explain the working of either horn type or cone type loudspeaker with the help of a neat diagram.
13. Why multispeaker systems are required?
14. Explain the difference between a subwoofer, woofer, mid-range speaker, supertweeter and tweeter.
15. What are the advantages of multispeaker system?
16. Explain the working of a simple crossover network.
17. With the help of a block diagram explain a hi-fi system.
18. Explain the following briefly.

a. Presence
b. Distortion
c. Input devices
d. Amplifier system

19. What is significance of a transformer in impedance matching?
20. How will you match a low impedance to a high impedance? Give a practical example.

Multiple Choice Questions

1. A loudspeaker converts
    a. electrical energy to mechanical energy
    b. acoustic energy to electrical energy
    c. mechanical energy to electrical energy
    d. electrical energy to acoustic energy
2. The input impedance of a crystal loudspeaker is
    a. resistive
    b. inductive
    c. capacitive
    d. a combination of (a) and (b)
3. Crystal speakers are used in
    a. low frequencies
    b. high frequencies
    c. mid frequencies
    d. both (c) and (d)
4. Electrostatic speakers need a
    a. dc polarising voltage
    b. ac polarising voltage
    c. either (a) or (b)
    d. neither (a) nor (b)
5. The dc bias (polarising voltage) in electrostatic speakers is of the order of
    a. 500 to 1000 V
    b. 1100 to 1500 V
    c. 1000 to 1200 V
    d. 500 to 1200 V
6. High class sound systems use electrostatics to reproduce
    a. low frequencies
    b. mid frequencies
    c. high frequencies
    d. both (a) and (b)
7. High quality dynamic speakers use diaphragms composed of
    a. paper
    b. titanium
    c. aluminium
    d. none of the above
8. Woofers are speakers designed to reproduce
    a. bass
    b. treble

c. neither (a) nor (b)
d. either (a) or (b)

9. Tweeters are speakers designed to reproduce
   a. low frequencies
   b. high frequencies
   c. mid frequencies
   d. both (a) and (c)
10. The smallest diaphragm for any half-way decent woofer is
   a. 4″
   b. 6″
   c. 8″
   d. 12″
11. High resonant frequencies are obtained with
   a. light weight
   b. stiff suspention
   c. loose suspension
   d. heavy weight
12. A horn-loaded unit
   a. increases the acoustic output
   b. decreases the acoustic output
   c. multiplies the acoustic output
   d. none of these

Fill in the Blanks

1. The principle of operation of a loudspeaker is the _______________ of that of a microphone.
2. Crystals have been known to _______________ under stresses.
3. To overcome frequency doubling we _______________ the speaker.
4. The voice coil is free to move _______________ but not vertically or laterally.
5. The stronger the magnet, the _______________ the speaker.
6. _______________ magnets have several advantages over those made of metal.
7. The _______________ is that part of the speaker which does the talking.
8. Mechanical inductance is called _______________.
9. Mechanical _______________ is called compliance.
10. Undesired sharpness of the resonant peak can be lessened by _______________.
11. Directivity normally varies considerably with _______________.
12. The _______________ does acoustically what the cone does mechanically.

ANSWERS

Multiple Choice Questions

1.	(d)
2.	(c)
3.	(b)
4.	(a)

5.	(c)
6.	(c)
7.	(a), (b) & (c)
8.	(a)
9.	(b)
10.	(c)
11.	(a) & (b)
12.	(a)

Fill in the Blanks

1.	reverse
2.	vibrate
3.	polarise
4.	back & forth
5.	better
6.	ceramic
7.	voice coil
8.	inertance
9.	capacitance
10.	damping
11.	frequency
12.	horn

CHAPTER 7

LOUDSPEAKER SYSTEMS

The end result of an audio system is to reproduce sound waves in space, with acoustic properties peculiar to each individual application. Hence, the choice of an appropriate loudspeaker system is of vital importance to the overall result. All that has been done in the electronic part of the system now relies on the loudspeaker system to successfully propagate the program into the space where it will be heard.

Every indoor installation must "live" with the acoustic characteristics of the room where it will operate. An outdoor installation is simpler, due to the absence of this problem, but it does have to withstand weather. Also, in general, it has to serve larger areas without help from reverberation—the containment of the sound.

HORNS

Outdoor speakers need the maxiumum efficiency, economic feasibility, and directivity so that sound is not wasted in directions where it is not needed or wanted. For these reasons, *horns have become almost universal for outdoor installations: they provide natural weather protection; they are relatively high in efficiency; and they have a degree of directionality dependent on the constants in their design.*

Horns come in all sizes and shapes, and are made of a variety of materials, metallic and plastic. Most of the smaller ones, suitable for paging, outdoor intercom, or low-level distribution, come with the driver unit as an *integral part.* The larger ones have *removable drivers* so that power can be increased, where necessary, by fitting a more powerful driver unit.

Metal horns are, in general, more rigid, but they do add a “tinny” quality to the reproduced sound. For many outdoor installations this is quite unimportant, as the sound is quite intelligible, which is the main requirement. Horns made of various *plastic substances* avoid the “tinny” sound, but may still give a trumpet-like effect, which is also quite acceptable in most outdoor installations. *The directivity pattern depends more on the internal design of the horn than on its shape, as observed externally*. A simple horn, with uniform development of the wave down its length, leading to a rectangular mouth opening, gives the widest dispersion across the narrower mouth dimension (Fig. 7.1). However, some directional horns are designed with changes inside the development that make it possible for the long dimension of the *mouth to correspond with the direction of wider dispersion*, which is psychologically what one expects.

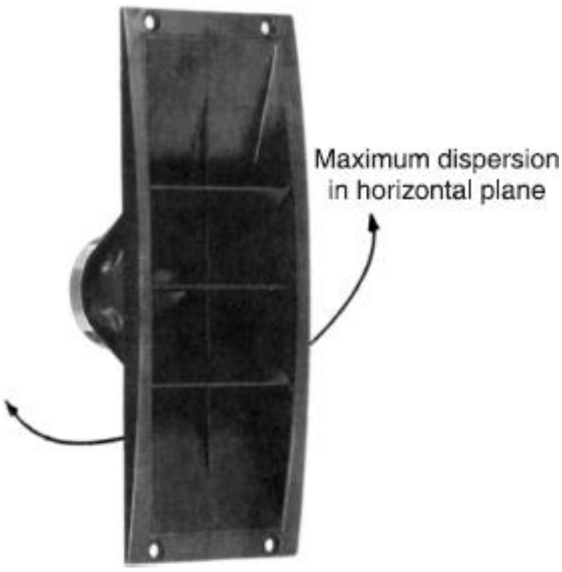


Fig. 7.1 A horn with a rectangular mouth set with its longer dimension vertically, provides wider dispersion in the horizontal plane. Horn size

Horn size relates not only to power-handling capacity: even more directly it relates to low-frequency cut-off. Every horn has a *low-frequency cut-off* below which it will not operate. If appreciable power at any frequency below this is fed to the unit, it may result in damage, and it will not radiate appreciable sound at such frequencies. *The smaller the horn, the higher the cut-off frequency*. A wise precaution in any system using horns is to install a *bass cut* that prevents excessive signal reaching the horns at frequencies below cut-off. The simplest method is a simple capacitor in series with the output, which may be applied, either one in each speaker connection, or one for the whole system (where all horns are used).

The value of the capacitor (Fig. 7.2) should be calculated to have a reactance equal to that of the load (the horn impedance or combined impedance) at the cut-off frequency. Thus if all the horns have a cut-off of 200 Hz, and the system consists of eight 16 W units, the combined impedance is 2 W. A reactance of 2 W at 200 Hz requires a capacitor of 400 mfd. Smaller horns may have a cut-off at a higher frequency, such as 800 Hz. If they are each of 45 W impedance, and a system uses 15 of them, the combined impedance is 3 W. A capacitor to have 3 W reactance at 800 Hz must be about 67 mfd. A 50 mfd capacitor of suitable working voltage should serve.

The problem of achieving adequate audience coverage in situations where background noise create problems is more likely to arise in outdoor than indoor situations. Buildings provide some insulation against outside noises getting in, and the confines of the building enable sound energy to be conserved, rather than escaping to infinity as happens outdoors.

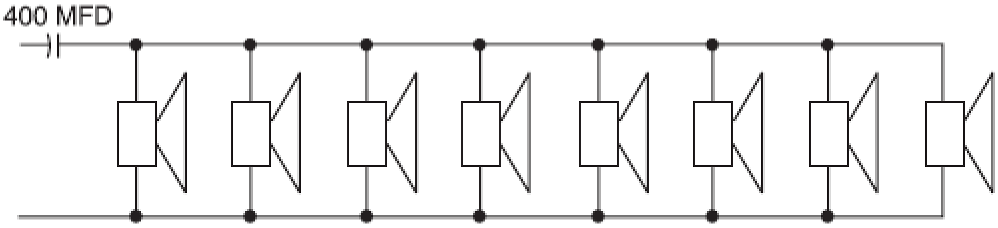


Fig. 7.2 The capacitor in this diagram acts as a simple bass cut-off to protect horns against receiving frequencies below cut-off.

INDOOR ACOUSTICS

The variety of situations that can be encountered indoors is virtually infinite. Studies in acoustics have established certain ideal characteristics, according to building size and the purpose or kind of sound that will make up program. At best these provide only *guidelines.*

The ultimate sound reproduced is a quite complex wave system, propagated through space. *In any indoor installation, every listener, or member of the audience, should hear first a direct sound at every component frequency, and then an appropriate amount of reverberation, or reinforcement of the same complex sound.* If there is no reverberative reinforcement the sound seems unnatural; if there is too much of the direct sound, it is not strong enough to be intelligible.

Reverberation “happens” whatever you do. Sometimes, as in studios, structural or decorative designs can be used to adjust its amount to better suit the purpose of the sound system. But more often, especially in home high fidelity and stereo installations, the owner just wants to put the system in his room—he's not about to redesign his room to suit the system. So we have to accept the room “*as is” and find a way of achieving a good reproduction illusion in that environment.*

In the living-rooms kind of environment, it seems as if nature is “on our side.” For smaller rooms the smaller speakers (such as the “bookshelf” variety) seem to do the better job, while for larger rooms the larger enclosures, such as corner horns, are better. But in case of larger installations there are more factors which affect the acoustics.

In the home system, room acoustics vary considerably; some rooms are carpeted, wall-to-wall; perhaps the windows are adorned with heavy drapes; and the condition is further augmented with an acoustic tile ceiling and plenty of well-stuffed furniture. Such a room is relatively dead acoustically. If you talk, the room gives the impression of being nice and quiet.

At the other end of the spectrum is the "empty" room, or one that sounds that way. The floor may be made up of tile or wood, the walls plaster or paneled, with a hard finish, and the ceiling completely nonabsorbent. The furniture may be devoid of upholstery, such as the wrought-iron frame type, with cloth draped over it like a sling. Such a room is *acoustically live.* If a number of people are holding simultaneous conversations in such a room, the effect can become quite nerve-wracking. For reproduced sound each kind of room needs its own treatment, for which different kinds of speaker systems are suited.

The reason why *speaker size should be suited to room size,* apart from the fact that the customer will usually prefer it that way, is that *speakers tend to generate sound waves commensurate with their physical size.* Of course, every frequency has its corresponding wavelength, and the speaker must have some way of developing waves of each wavelength in the audio spectrum. Large speakers require more space to develop the waves, particularly the lower frequency ones, than the smaller speakers. Consequently, placing a larger speaker in a relatively small room may not give the waves space to develop as the speaker designer intended. This is particularly true for corner horns, and to a lesser degree of bass reflex types.

STEREO SYSTEMS

As most home high fidelity systems are not stereo, we will devote this section directly to achieving the best stereo illusion in different type of rooms. For the *medium-to-dead type of room* with plenty of absorptive surfaces, the conventional stereo placement is best: two speakers are spaced apart so that their separate program content can be clearly heard through most of the room.

In a *conventional rectangular room* the speakers should be placed either in the corners at opposite extremities of one of the shorter walls, or a little way in from the corners along one of the longer walls (Fig. 7.3). Speakers types can be chosen to suit the situation. In a bigger room, large corner speakers may be best. In a smaller one, perhaps suitable placement for the "bookshelf" type can be found. But even in this type of room, sometimes the ideal placements are not practical. Not all rooms are rectangular the *L–shaped room is* very popular — and doors, windows, and fireplaces are often right where one of the speakers ought to be.

Perhaps an aid to understanding what we try to do in stereo is the concept of an *axis of symmetry*. In case of speakers at or near the extremities of one wall, the axis of symmetry bisects that wall and extends across the room, dividing it into two similar halves. In the case of a square room another perfect axis of symmetry would be a diagonal (Fig. 7.4). But in a rectangular room an approximate axis of symmetry could run from corner to corner, a similar diagonal. Putting the speakers on appropriate mid-sides gives a whole range of new possibilities for *stereo placement.*

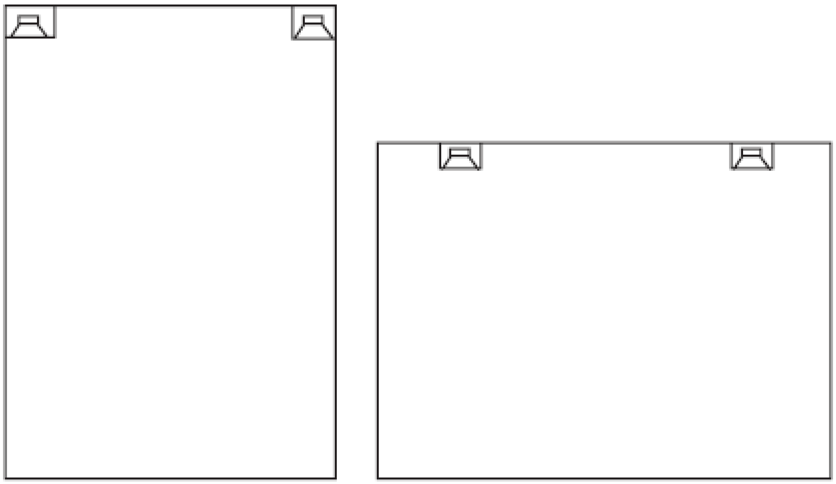


Fig. 7.3 This sketch shows ideal speaker placements in a rectangular room with fairly “dead” acoustic properties.

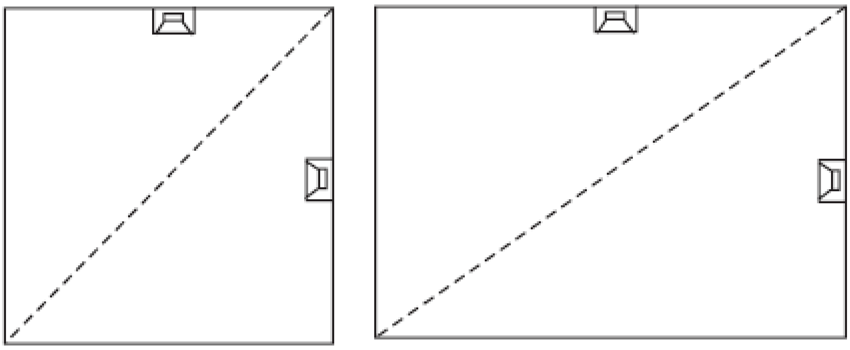


Fig. 7.4 A diagonal axis of symmetry may be used to determine stereo speaker placement; left in a square room, right in a rectangular one.

Rooms with L shapes or other deviations from the simple rectangular, pose a different kind of problem. But usually the simplest concept is to think of the shape as a modified rectangular. The basic rectangle, on which you base your notion of providing stereo, may be part of the whole room, or more than the whole room. For example, if the dining area is a relatively small area added on to the living room, and people will not normally listen to stereo while dining, or will only treat it as incidental music if they do, the best approach is to ignore the existence of the dining area (Fig. 7.5). The fact that some of the sound from one side, more than the other of the stereo, will spill over into the dining area may require a slight *balance adjustment* of the stereo controls, so the balance seems correct.

On the other hand, some L-shaped rooms could be more accurately approximated as a larger rectangle from which a small area is removed (Fig. 7.6). This is particularly true where the remaining area happens to be kitchen, which is only separated from the living-dining area by a counter-top, rather than a complete wall (Fig. 7.7). In either of these cases, it would be easier to consider that you are aiming to serve the entire, large rectangular area with stereo.

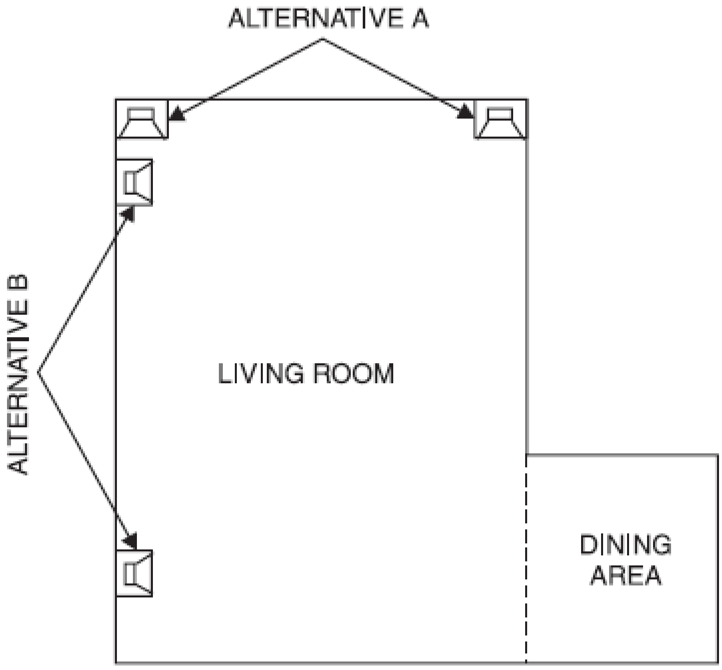


Fig. 7.5 Where the dining area is small in an L-shaped room, its existence may be ignored in considering speaker placement for the living room area by adjusting the electrical balance to compensate for sound absorption by the dining area.

Where the room is far more *live*, as in the average recreation room, without carpets, drapes (heavy ones at least), or acoustic tile ceiling, or with only some of these, conventional treatment as suggested above may not give a very good stereo illusion because there is too much echo. In the dead rooms the aim is to serve the entire listening area with sound direct from both loudspeakers (with the possible exception of spill over areas). If sound doesn't reach the listener directly, it's apt to be lost, so sitting in a position where only one speaker can be heard will lose the stereo illusion. But in the live room, reflections tend to destroy the illusion if this approach is used. So *the technique in a live room is to utilise the reflections, since you cannot easily get rid of them.* One way to do this is to use a cabinet-type stereo with speakers on the ends of the cabinet (Fig. 7.8). This relies on reflections from the walls to get the apparent separation, rather than allowing the sound bounced off the walls to destroy it.

Another technique that works well in this kind of room, although it is not confined to this type, is the so-called dipole or planer-type speaker. All the speakers we have discussed to this point radiate what is basically a sound pressure wave, and the point where this wave enters the room, (or a reflection of it, in the case of Fig. 7.8) identifies the sound source for each channel. The *dipole speaker* doesn't work in the same way. The back is completely open so that when the front pushes a sound **compression** out, the back sucks a sound rarefaction in. Listening to one of the speakers on monophonic program, as you walk around it, the back and front do not sound too different from conventional speakers. But when you get edge on, the location of the speaker suddenly seems to vanish (Fig. 7.9). You suddenly get the impression that the thing you're looking on as a speaker isn't working, and the sound must really be coming from somewhere else.

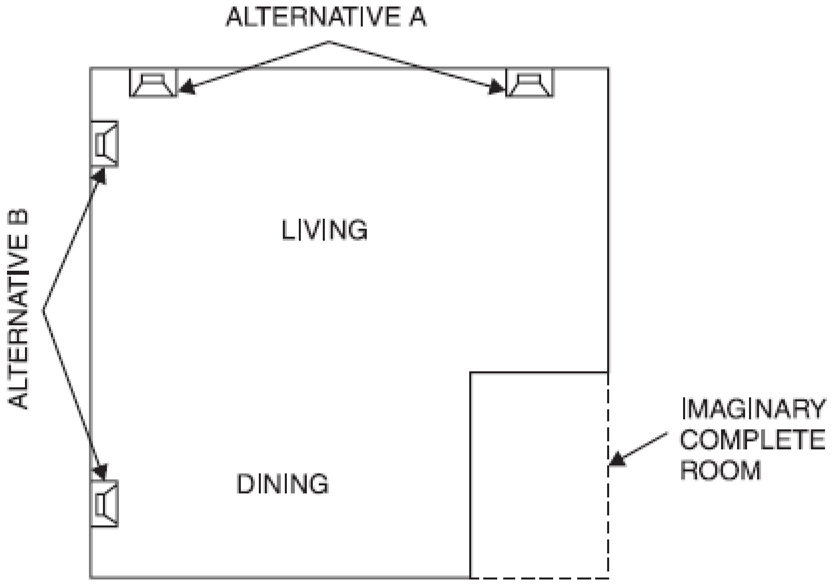


Fig. 7.6 Some L-shaped rooms approximate more closely a large room with a piece missing. In this case the placement is aimed at covering the whole of the larger, theoretical area.

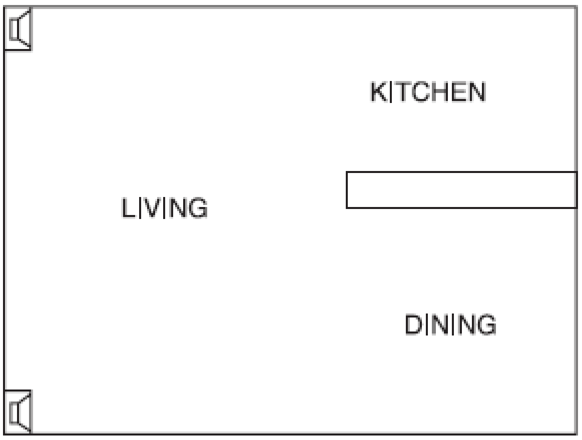


Fig. 7.7 Where kitchen-dining areas have relatively open access between them, the effective room to be served should include the whole of the larger rectangle.

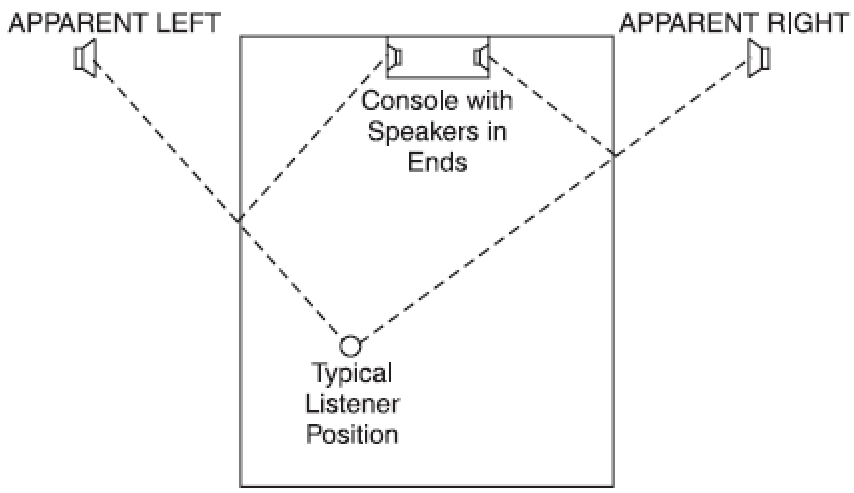


Fig. 7.8 A console type package system provides good stereo in a highly reflective, "live" room, with the speakers in the ends facing out. This utilises the reflections (rather than treating them as "unwanted") to achieve a greater apparent separation.

If two of these are used on stereo, one for each channel, and placed in a manner somewhat like that shown in Fig. 7.10, a completely new type of listening situation results. Now, in either end of the room, the stereo illusion is as good as it would be with the other type, similarly placed, but facing only toward you. If you used two for each channel, back-to-back, in a highly reflective room, you'd get too much reflection from the ones facing away from you via the far end of the room. But with dipole speakers placed like that in Fig. 7.10, not only do you get a good stereo illusion in both ends of the room, it is also good immediately between the speakers, although you would not identify the sources of sound with the speakers. However, you do get the *separation*, because of the interaction between the different program content in the two channels.

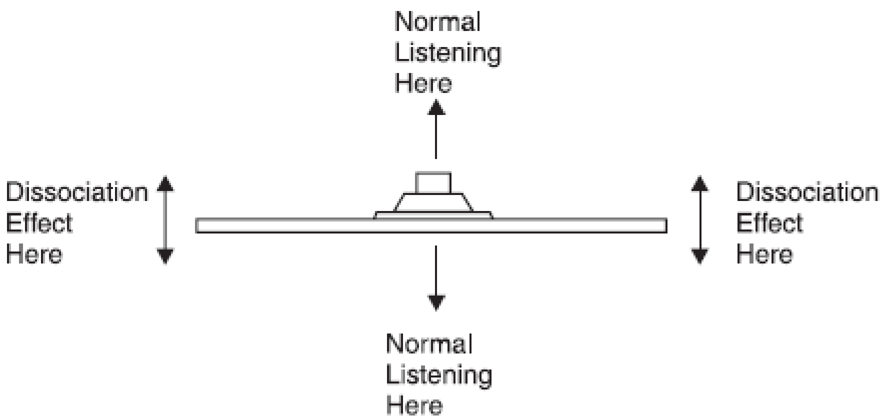


Fig. 7.9 Illustrated in this sketch is another kind of loudspeaker, which gives "dipole" radiation of sound. Used on single-channel sound, the listening effect varies with position.

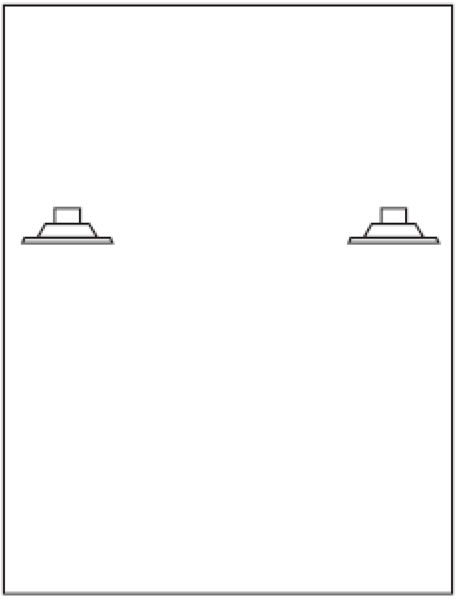


Fig. 7.10 This stereo situation empolys two open backed or dipole-radiator type loudspeakers.

COST AND EFFICIENCY

The general methods of providing stereo for different kinds of rooms are already considered, but another important aspect which needs mention is *cost and quality.* And this in turn is related to *efficiency*, especially when size is taken into consideration. Small speakers can be highly efficient, provided their frequency response isn't very low, since as small speakers become more effective at lower frequencies (bass) they lose efficiency.

Earlier books on this subject related quality and efficiency to magnet size—which is the most costly single part of a loudspeaker. However, with the newer ceramic magnets, this relationship is not quite so certain, because the magnet is no longer the vital cost element. And even if magnet size is considered an index of efficiency, this is true only up to a certain size, when other limitations stop a further rise in efficiency with increasing magnet size.

Obviously, if magnet size were the only factor in determining efficiency, this could rise no higher than 100%. After a magnet size sufficient to achieve 50% is reached, further improvement in efficiency would not be commensurate with the increase in magnet size and cost needed to achieve it. But when an objective, such as packing a speaker with good bass response into a small box, is taken into account, further basic *limitation* on efficiency is imposed, so that the theoretical ultimate might be 2% instead of 100% (which isn't possible, in fact, because there is no solid substance as light as air which can be used to construct a diaphragm). Taking all these factors into account, *a reasonable "high-efficiency" speaker—one of the larger variety—might run about 25% efficient. A low-efficiency speaker, one of the bookshelf variety, is more likely to run about 2% efficient.*

Efficiency is not necessarily related to quality. This is a matter of how well the individual design is implemented. Usually, within the other limitations, improving efficiency within a given type goes along with improved quality, and also higher cost. But technology is advancing so rapidly that a low-cost, well-designed speaker of recent manufacture may be better than older ones costing far more, and on which possibly more design hours were spent!

Speakers for *indoor applications*, other than high fidelity or stereo in the home, are made to more competitive economic standards. Unless difficult conditions make it essential to get the maximum acoustic power, requiring a high electrical power with highly efficient speakers, using speakers slightly less efficient can save considerably on cost, and a little more electrical power can be obtained more cheaply than the cost of upgrading all the speakers in the system. Speakers for paging and intercom systems are usually smaller than for the same environment in other applications, with a restricted frequency range and higher efficiency.

MULTIWAY SYSTEMS

Interestingly there are two philosophies in overall speaker design, one which aims to make the unit as efficient as possible, with the argument that a high efficiency maintained over the frequency range must have a fairly uniform response, by the very nature of it; and the other which sacrifices efficiency for a reduction in size. So there are two philosophies about covering the entire audio frequency range.

Many multiway system designers like the superior efficiency of horns. *A horn is actually a matching device that works like a transformer to match the moving diaphragm to the ultimate air wave formed at the mouth of the horn, which is not possible in cone type.* **So, over the range for which it is effective a horn speaker is more efficient, and can give a smoother response than the cone type.**

It should not be assumed that the horn type is better, per se. Achieving this ideal requires careful design and close precision production. Possibly, equal attention to these aspects can achieve at least equal performance from the cone-type speaker. But audio people tend to be idealists and, if they believe the horn is inherently better, they go that way, regardless.

A disadvantage of the horn type is that *when it fails to maintain its uniformity, it does so relatively suddenly.* At the low-frequency end a horn has a cut-off frequency below which the horn quite abruptly fails to function as a horn. And at the high-frequency end any factors at the throat or in the horn development that cause deterioration there usually do so quite emphatically, with serious dips and peaks in the resulting response. So a multiway system, Figs. 7.11 and 7.12 using one or more horns (some systems combine cone units for some frequency ranges with horn speakers for others) must employ relatively sharp crossover filters so that very little energy is applied to the horn beyond its useful range of frequencies. Crossover filters should be complementary, or the overall response will not be correct. So making the filters sharp for the horns means that filters used for accompanying cone-type units must be equally sharp.

There are two disadvantages to high-slope crossovers: one is electrical and the other acoustical. *Electrically*, the sharper crossovers are more dependent for precisely correct response and on correct termination than are the simpler types. With the electronic crossover, of which few, if any, high-slope versions are currently available, this can be accurately built into the unit with the use of selected precision components. With the electrical crossover applied at the output, between a single power amplifier and multiple speaker units, the filters are terminated by speaker unit impedances, which are not simple resistances. So the simpler crossovers more easily achieve the responses they are intended to give.

Acoustically, the effect of phase response can be important. Tests have been made with subjective listening that suggest human hearing is not sensitive to phase differences so long as relative levels are not seriously modified (such as by cancellation). There is a fairly obvious reason for this: acoustic environment produces quite a variety of phase shifts, but the basic character of the sound is still interpreted, in spite of the environment, provided this is not severe enough to cause what may be recognised as colouration.

But the phase shifts thus considered are fairly small. If the changes are equivalent to sending the sound through a constricted, resonant pipe, it sounds like it. In a network that provides an ultimate roll-off of 24 dB/octave, the phase difference between the two outputs at every frequency (which is most important at crossover) is 360°, or a complete wave. Although the two waves are in phase with each other at that frequency, they are a whole period apart. And in any composite signal most of that 360° phase shift of a whole wave occurs between a fundamental and its harmonics. And if the system is more than two-way the effect of the various crossovers is additive. A sweep through the frequency range with a sinusoidal tone may give the impression that the response is extremely uniform–as it may well be—but the handling of this system applied to transient program signals may be another thing altogether.

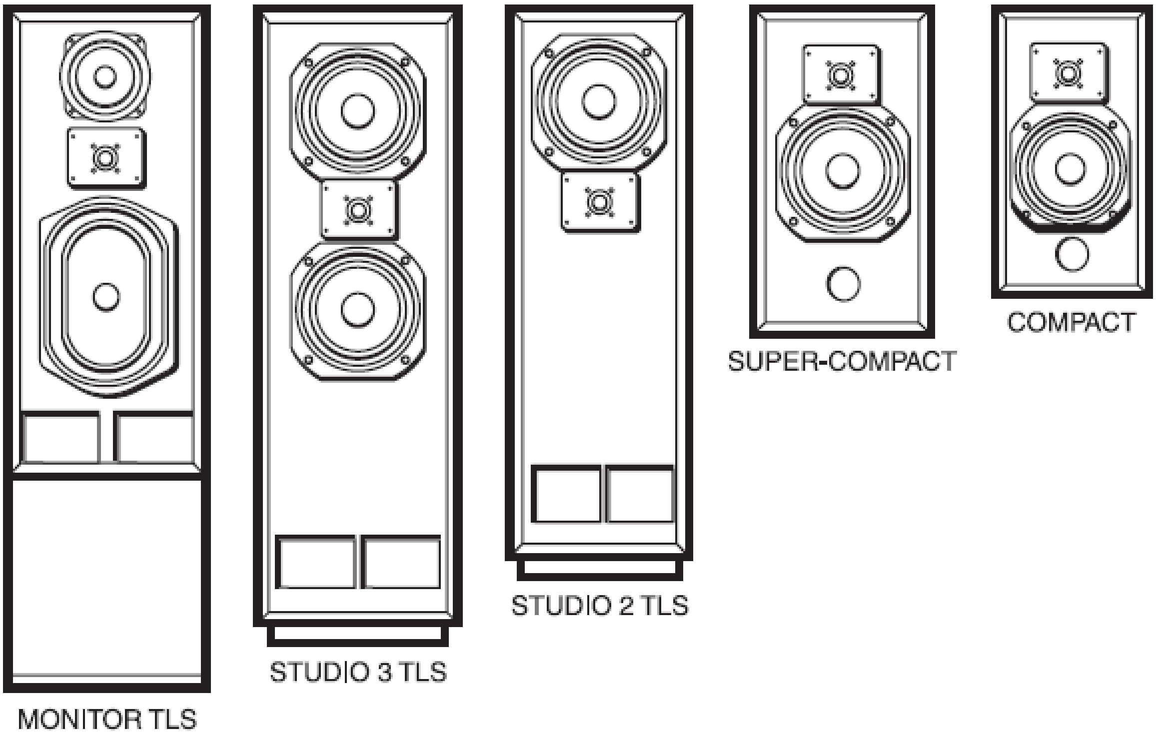


Fig. 7.11 Standard transmission line speakers

Fig. 7.12 Linear phase speaker systems

These comments applied to multiway systems on a frequency division basis do not apply to systems that use a number of similar units to achieve the same or different objectives. For example, a group of units that would not handle the lower frequencies individually may do so easily when they are mounted on a common baffle board, or in the same cabinet, and driven in unison. Consequently, *a group of units will handle a wider range of frequencies than each unit can individually. And the beam-type unit, which employs a row of units to make a line radiator, applies a combination of units to a different objective—the direction of sound where it is wanted and avoiding sending it where it is not wanted.*

Fig. 7.13 Supersonic speaker system

COMPONENT SYSTEMS ENHANCE YOUR HIGHS

In most cars, speakers mount at the bottom of the door. That's fine for bass, but it doesn't do much for the high frequencies.

Component speaker systems, with separate midrange/woofers and tweeters, let you put the bass down low and the highs where they belong—near your ears. You can mount the tweeters at the top of your door or on the dash, filling your car with a breathtaking stereo effect.

Component systems come with true crossover networks. The crossovers divide the sound between the tweeter and the mid/woofer, taking full advantage of the strengths of each component. Some models include tweeter level controls so you can fine tune the sound to suit you ear.

EXERCISES

Descriptive Questions

1. Differentiate between the acoustical characteristics of an indoor installation and an outdoor installation.
2. How will you achieve the best stereo illusion in different type rooms?
3. Explain the significance of a multiway speaker system.
4. The speaker size should be suited to room size. Explain.

Fill in the Blanks

1. Every indoor installation must live with the ________________ characteristics of the room where it will operate.
2. An outdoor installation does have to withstand ______________.
3. Outdoor speakers need the ____________________ efficiency economically feasible.
4. Horns have become almost universal for ________________ installations.

5. Horns provide natural weather ___________________.
6. Horns are relatively __________________ in efficiency.
7. Smaller horns come with the ___________________ unit as an integral part.
8. Larger horns have ___________________ drivers so that the power can be increased, where necessary, by fitting a more ___________________ driving unit.
9. The directivity pattern depends more on the ____________________ design of the horn than on its shape, as observed externally.
10. Horn size relates to __________________ handling capacity and ____________________ frequency cutoff.
11. In any indoor installation every listener should hear first a ____________________ sound at every component frequency, and then an appropriate amount of ___________________ of the same complex sound.
12. Speakers tend to generate sound waves _____________________ with their size.
13. The technique in a live room is to utilise the ___________________, since you cannot easily get rid of them.
14. A resonable high-efficiency speaker might run about _________________ % efficient.
15. A low-efficiency speaker is more likely to run about _________________ % efficient.

ANSWERS

Fill in the Blanks

1.	acoustic
2.	weather
3.	maximum
4.	outdoor
5.	protection
6.	high
7.	drive
8.	removable, powerful
9.	internal
10.	power, low
11.	direct, reverberation
12.	commensurate
13.	reflections
14.	25
15.	2

UNIT 2

The word television means viewing from a distance. In a broader sense, television is the transmission of picture information over an electric communication channel. It is desired that the picture picked up by television receiver be a faithful reproduction of the scene televised from a TV studio. Any scene being televised has a wide range of specific features and qualities. It may contain a great number of colours, gradations of shade, coarse and fine details; motion may be present in a variety of forms and the objects making up the scene are usually in three dimensions.

In the early 20th century, TV systems, primitive from the standpoint of the present day, were mechanical. Today, TV equipment is all electronic. The replacement of mechanical by electronic television has made it possible to reproduce a high-quality picture approaching the original scene in quality.

ELEMENTS OF A TELEVISION SYSTEM

Television is an extension of the science of radio communications, embodying all of its fundamental principles and possessing all of its complexities and, in addition, making use of most of the known techniques of electronic circuitry.

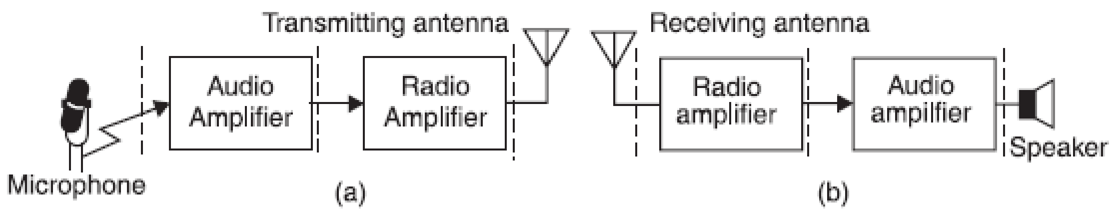


Fig. 30.1 Simplified block diagram of a (a) sound transmitter and (b) a sound receiver

In the case of the *transmission and reproduction of sound*, the fundamental problem is to convert *time variations* of acoustical energy into electrical information, translate this into radio frequency (RF) energy in the form of electromagnetic waves radiated into space and, at some receiving point, *reconvert* part of the resultant electromagnetic energy existing at that point into acoustical energy.

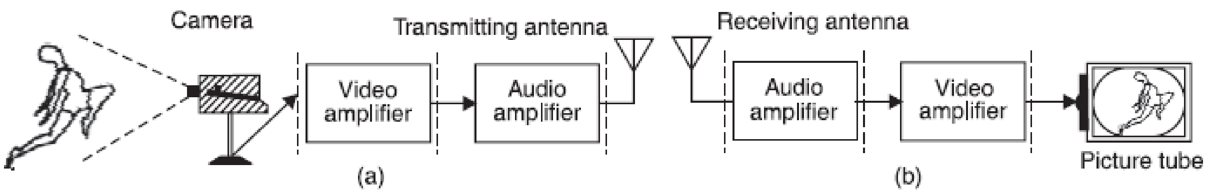


Fig. 30.2 Simplified block diagram of (a) picture transmitter and (b) picture receiver

In the case of *television*, there is the parallel problem of converting *space and time variations* of luminosity into electrical information, transmitting and receiving this, as in the case of sound, and *reconverting* the electrical information obtained at a receiving location into an optical image.

When the information to be reproduced is *optical* in character, the problem fundamentally is much more complex than it is in the case of *aural* information. In the latter instance, this is dealt with at each instant of time only a single piece of information, since any electrical waveform representing any type of sound is a *single-valued function of time* regardless of the complexity of the waveform. In the corresponding optical case, at any instance there is an *infinite number of pieces of information existing simultaneously* namely, the brightness which exists at each point of the scene to be reproduced. In other words, the information is a *function of two variables, time and space.* Since the practical difficulties of transmitting all this information simultaneously and decoding it at the receiving end at the present time seem insurmountable, *some means must be found whereby this information may be expressed within the form of a single-valued function of time.* In this conversion, the process known as *scanning* plays a fundamental part.

THE SCANNING PROCESS

Scanning may be defined as *the process which permits the conversion of information expressed in space and time coordinates into time variations only.* Suppose, for extreme simplicity, that an optical image of a scene, perhaps on a photosensitive surface, is scanned by a beam of electrons, i.e. all points on the image are *sequentially* contacted by this beam, and that somehow, as a result of this scanning, *through capacitive, resistive, or photoemissive effects at the surface*, an electrical signal may be obtained that is directly proportional in amplitude to the brightness at the particular point being scanned by the beam. Although the picture content of the scene may be changing with time, if the scanning beam moves at such a rate that any portion of the scene content does not have time to move perceptibly in the time required for one complete scan of the image, the resultant electrical information contains the true information existing in the picture during the time of the scan. This *derived information* is now in the form of a signal varying with time.

Consider the image of Fig. 30.3 (a). The light and dark areas represent *variations in brightness* of the original scene. Suppose that a beam of electrons is made to scan this image as shown, starting in the upper left hand corner and moving rapidly across the image in a time t, thus forming line 1 as shown, and then made to return instantaneously. This process is repeated until the bottom of the image is reached. The *time variation* of the horizontal component of motion of the scanning beam is that shown in Fig. 30.3 (b).

Means exist whereby an electrical signal corresponding in amplitude to the illumination on the point being scanned may be derived. The electrical output as a function of time corresponding to individual scanning lines is shown in Fig. 30.3 (c). This electrical information corresponding to the light intensity of a large sampling of points in the original scene is now obtained, with the limitation that *the detail which can be reproduced depends on the completeness of coverage by the scanning beam.* This coverage is determined directly by the total number of scanning lines. If the scanning beam is now made to return to the top of the picture and this process is *repeated*, another sampling of information identical with the first will be obtained unless the scene has changed in the meantime, in which case the new sampling of information will be in accordance with this change.

At the receiving end, the fundamental problem is that of *recombining* the information, which has been broken down into a time function, into an optical image. The image which is reproduced, with the number of lines shown, is indicated in Fig. 30.3 (d).

SCANNING METHODS AND ASPECT RATIO

The *scanning mechanism* is the portion of the entire television system which deserves the most attention from the standpoint of the necessity of *formulation of standards of methods and performance.* Once the scanning system is standardised, the performance of the system is specifically limited.

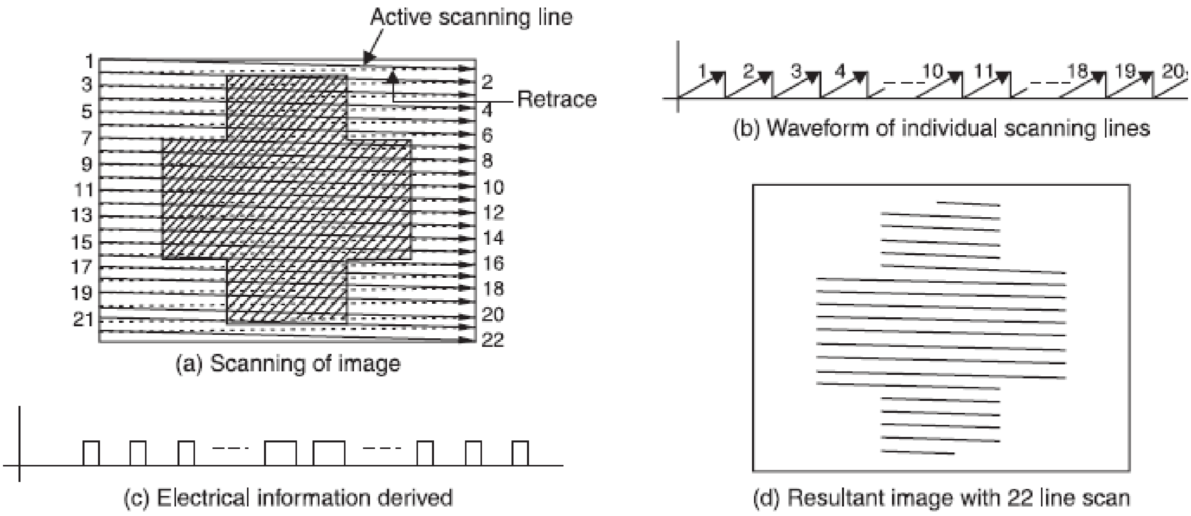


Fig. 30.3 The scanning process

The process of scanning makes possible the use of a *single transmission channel.* Otherwise there would be required as many transmission channels as there are simultaneous units of optical information to be reproduced.

In the scanning process the amount of detail actually converted to useful information depends on *the total percentage of picture area actually contacted by the electron beam.* The scanning process is such that the picture area is traversed at repeating intervals, giving in effect *a series of single pictures* much in the same manner that motion pictures are presented. The *repetition rate* of these successive pictures (referred to as *frame rate)* determines the apparent *continuity* of a moving scene. This rate must be chosen sufficiently high so that neither *discontinuity of motion* nor *flicker* is apparent to the eye.

The scanning method chosen might depend on the *geometry* of the image to be reproduced. It is logical that the geometry should be chosen by the viewer of the final reproduced picture. Here it is useful to rely on the experience of the motion-picture industry, which until the advent of stereo techniques reproduced a *rectangular pattern* having an *aspect ratio* (picture width to picture height) of approximately 4 : 3. Most subjective tests have indicated that *aspect ratios approximating 4: 3 are most pleasing artistically and less fatiguing to the eye.* A physiological basis for this might be that the eye is less restricted in its range of movement in a horizontal than in a vertical direction. Also, the fovea, or area of greatest resolution, is some what wider than it is high. Thus, *the area of the fovea is most efficiently utilised.*

Once the geometry of the image has been specified, it remains to determine what *sequence* may be used by the scanning beam to cover the entire area. For a picture of rectangular geometry, any form of spiral or circular scanning would be wasteful of scanned area. Also, scanning of this type would result in *nonuniform coverage* of the area or in varying scanned velocities. A varying scanning velocity for the pick-up and reproducing devices results in varying output for constant scene brightness. Correction for this would be necessary.

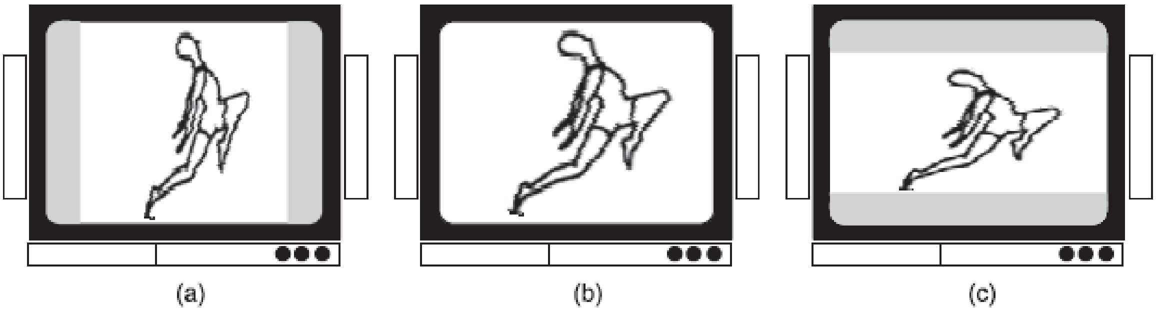


Fig. 30.4 Aspect ratio (a) A picture in which the width-to-height proportion is less than 4:3. (b) A picture in which the width-to-height proportion is NORMAL. Width is to height as 4:3. (c) A picture in which width to height proportion is greater than 4:3.

Most practical television systems, therefore, resort to *linear scanning* of the image. Also, *it is common practice to scan the image from left to right, starting at the top and tracing successive lines until the bottom of the picture is reached, then returning the beam to the top and repeating the process.* This is shown in Fig. 30.5 (a). The direction of the arrows on the heavy lines indicates the forward *(trace)* or useful scanning time. The dashed lines indicate the *retrace* time, which is not utilised in converting picture information into useful electrical information. The movement during the return of the scanning spot to the top of the image is shown in Fig. 30.5(b).

This is the type of scan which would result in a cathode ray device when the scanning-spot position is determined by the horizontal and vertical components of electric or magnetic fields where these components are *repeating linear functions of time* as shown in Fig. 30.5 (c). If the repetition rate of the horizontal component is related to that of vertical component by a factor n, then n lines are formed during a complete vertical period. *The retrace times, both horizontal and vertical, are not utilised for transmitting a video signal but may be employed for the transmission of auxiliary information* such as teletext, for example.

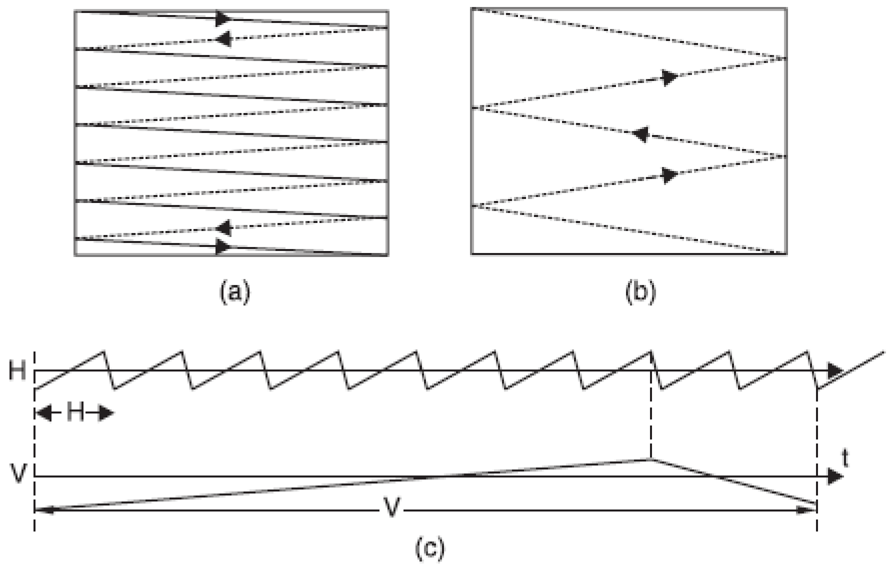


Fig. 30.5 Linear Scanning (a) Top to bottom scanning path of the beam (b) Vertical retrace path of the beam (c) Deflecting field waveforms

PERSISTENCE OF VISION AND FLICKER

The image formed on the retina is *retained* for about 20 ms even after optical excitation has ceased. This property of the eye is called *persistence of vision*, an essential factor in cinematography and TV for obtaining the illusion of continuity by means of rapidly flashing picture frames. If the flashing is fast enough, the *flicker* is not observed and the flashes appear continuous. The repetition rate of flashes at and above which the flicker effect disappears is called the *critical flicker frequency* (CFF). This is dependent on the brightness level and the colour spectrum of the light source.

In cinema, a film speed of 16 frames per second was used in earlier films to obtain the *illusion of movement.* Lack of smooth movement was noticeable in these films. The present day standard for movie film speed is 24 frames per second and at this speed, flicker effects are very much reduced. This problem is further reduced in modern projectors by causing each frame to be illuminated *twice* during the interval it is shown by means of fan blades. The resulting flicker rate is quite acceptable for cine screen projection, because it is viewed in subdued light and a wide display area.

In television, the field rate is concerned with (i) large area flicker, (ii) smoothness of motion, and (iii) motion blur in the reproduced picture. As the field rate is increased, these parameters show improvements but tend to saturate beyond 60 Hz. Further increase does not pay off, and increases the bandwidth. Hence the picture field scanning is generally done at the same rate as the mains power supply frequency, which conveniently happens to be 50 or 60 Hz for the same reasons of reducing illumination flicker from electric lamps. At 60 Hz the flicker is practically absent. At 50 Hz, a certain amount of *borderline flicker* may be noticed at high brightness levels used to overcome surrounding ambient light conditions. Use of commercial power frequency for vertical scanning also reduces possible effects, like supply ripple and 50 Hz magnetic fields, in the reproduced picture.

VERTICAL RESOLUTION

An electron beam scanning across a photosensitive surface of a pick-up device can be made to produce in an external circuit a time-varying electrical signal whose amplitude is at all times *proportional* to the illumination on the surface directly under the scanning beam. The upper frequency limit of the amplifier for such signals depends upon the *velocity of the scanning beam.*

Most scenes have brightness gradations in a vertical direction. The ability of a scanning beam to allow reproduction of electrical signals corresponding to these gradations may be seen to be dependent upon the *number of scanning lines* and not upon the velocity of the individual scanning lines. However, the two are integrally related if the picture-repetition rate and the number of lines are fixed.

It is possible to arrive at some estimates of the number of lines necessary to reproduce known variations in light intensity in a vertical direction from consideration of the *line pattern.* in Fig. 30.6.

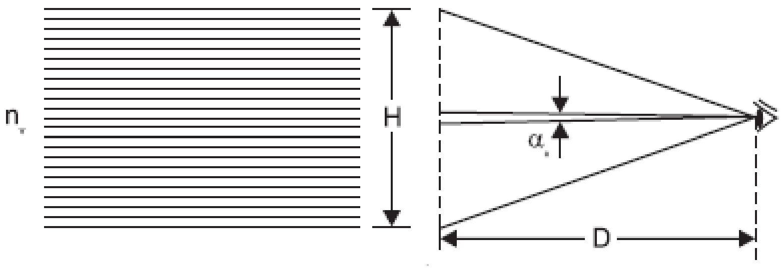


Fig. 30.6 Determination of the maximum number of black and white lines for vertical resolution

The scanning lines used per frame is referred to as the *scan ratio* of a television system. The realistic limit to the number of lines is set by the *resolving capability* of the human eye, viz. about one minute of visual angle. For comfortable viewing an angle of about 10 to 15° can be taken as the *optimum visual angle.* Hence the best *viewing distance* for watching television is about 4 to 8 times the height of the picture, i.e. a visual angle of about 10° as shown in Fig. 30.6.

The maximum number of dark and white elements which can be resolved by the human eye in the vertical direction in a screen of height H is given by n_v according to the relation:

$$H/D = n_v \times \alpha_o \qquad \text{...(30.1)}$$

where n_v is the number of black and white lines of vertical resolution, α_o is the minimum angle of resolution in radians and D the distance of the viewer.

Problem 30.1 Calculate the number of black and white lines of vertical resolution for a visual angle of about 10°.

Solution When α_o equals 10°, D/H = 6

$$n_v = \frac{H}{D\,\alpha_0} = \frac{1}{6}\times\frac{60}{1}\times\frac{180}{\pi} = \textbf{600 lines}$$

Problem 30.2 Calculate the number of black and white lines of vertical resolution for a visual angle of about 15°.

Solution When α_o equals 15° D/H = 4

$$n_v = \frac{H}{D\,\alpha_o} = \frac{1}{4}\times\frac{60}{1}\times\frac{180}{\pi} = \textbf{900 lines}$$

The number of alternate black and white elementary horizontal lines which can be resolved by the eye are thus 600 for 10° visual angle and 900 for a 15° visual angle. There has been a large difference of opinion about what constitutes the maximum number of lines because of the subjective assessment involved. This has been made further difficult to assess because of the effects of the *finite size* of the scanning beam spot.

PICTURE ELEMENTS

A picture element (pixel) is the smallest area in an image that can be reproduced by the video signal. The size of a *pixel* depends upon the size of the scanning spot, the number of scanning lines and the highest frequency utilised in the video signal. Its *vertical dimension* is equal to the distance between two scanning units, or centre to centre distance between two adjacent spots. Its *horizontal dimension* will also be the same.

With an aspect ratio of 4:3 the number of pixels on a horizontal scanning line will be 4/3 times the number of pixels on a vertical line. *The total number of pixels activated on the screen reproduce the picture.*

THE KELL FACTOR

In practical scanning systems, the maximum vertical resolution realised is less than the actual number of lines available for scanning. This is because of the finite beam size and its *alignment* not coinciding with the elementary resolution lines.

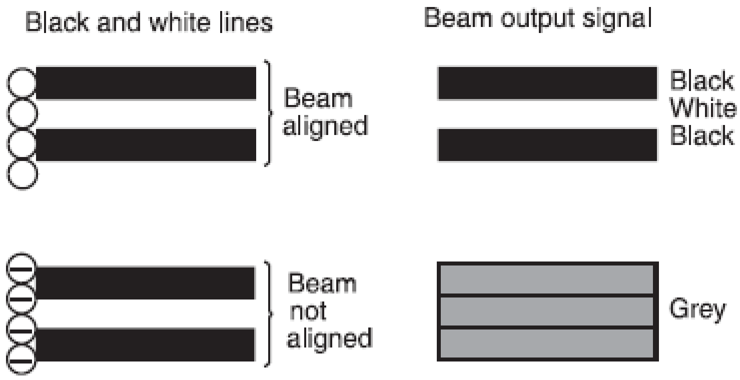


Fig. 30.7 Effect of beam spot size on vertical resolution

Consider a finite size of a beam spot scanning a series of closely spaced horizontal black and white lines of minimum resolvable thickness, when the beam spot size is compatible with the thickness of the line, as shown in Fig. 30.7. If the beam is in *perfect alignment* the output will exactly follow the lines as *black or white levels.* If, however the beam spot is *misaligned*, it senses both black and white areas simultaneously. Hence, it integrates the effects of both areas to give a resultant *grey level* output in between the black and white levels. This happens for all scanning positions and the output as well as the reproduced picture is a continuous grey without any vertical resolution at all. In positions of intermediate alignment of beam, it will be more on one line than on the adjacent line and the output will be reproduced with *diminished contrast.* This indicates that *there is a degradation in vertical resolution due to finite beam size.*

The factor indicating the reduction in effective number of lines is called the *Kell factor*. It is not a precisely determined quantity and its value varies from 0.64 to 0.85. What the Kell factor indicates is that *it is unrealistic to consider that the vertical resolution is equal to the number of active lines.*

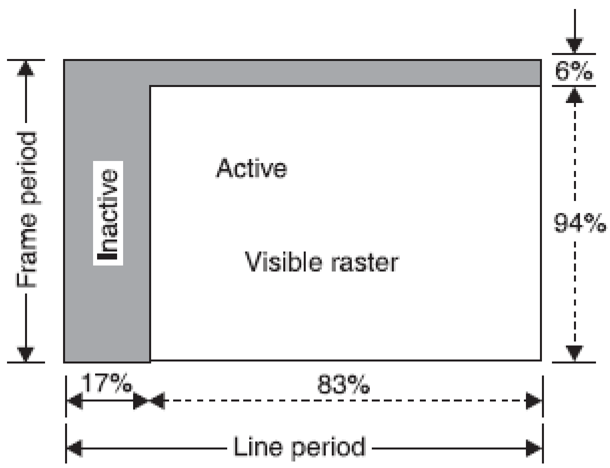


Fig. 30.8 Active and inactive scanning periods

The flyback time also reduces the number of active picture elements as depicted in Fig. 30.8. The *horizontal flyback interval* occupies 17% of the line period and *the vertical flyback interval* occupies 6% of the frame period.

In the 625 line system, the number of active lines lost in vertical blanking (referred to as *inactive lines)* are (625 – 40 =) 585 lines. With a Kell factor of 0.7 (average value) the vertical resolution is (0.7 × 585) = 409.50 lines. The *horizontal resolution may not exceed this value multiplied by the aspect ratio.*

HORIZONTAL RESOLUTION AND VIDEO BANDWIDTH

The horizontal resolution of a television system is the ability of the scanning system to resolve the horizontal details i.e., *changes in brightness levels of elements along a horizontal scanning line.* Since such changes represent vertical edges of picture detail, it follows that horizontal resolution can be expressed resolution as a measure of the ability to reproduce vertical information.

The horizontal resolution in a scanning system depends on the rate at which the scanning spot is able to change brightness level as it passes through a horizontal line across the vertical lines of resolution shown in Fig. 30.9.

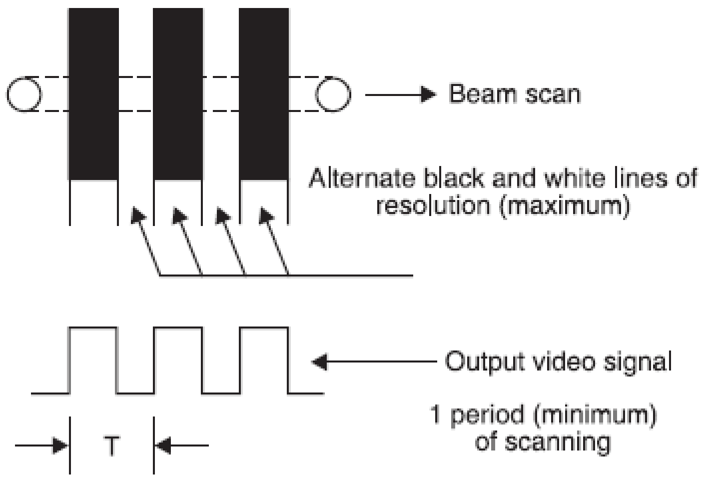


Fig. 30.9 Horizontal resolution and video bandwidth requirements

In the 625 line system, there are about 410 active lines of resolution. With an aspect ratio of 4:3 *the number of vertical lines for equivalent horizontal resolution* will be (410 × 4/3 =) 546 black and white alternate lines which corresponds to (546 × 1/2 =) 273 cycles of black and white alternations of elementary areas. For the 625 line system, the horizontal scan frequency (line frequency) is given by:

$$f_H = \text{number of lines per picture} \times \text{picture scan rate}$$
$$= 625 \times 25 = 15,625 \text{ Hz} \qquad ...(30.2)$$

as each picture line is scanned 25 times per second. The *total line period* is thus

$$T_H = \frac{1}{f_H} = \frac{1}{15,625}\text{sec} = 64\ \mu\text{s} \qquad ...(30.3)$$

Out of this period, 12 µs are used for the *blanking of flyback retrace.* Thus 546 black and white alternations (273 cycles of complete square waves) are scanned along a horizontal line during the forward scan time of (64 – 12 =) 52 µs. The period corresponding to this square wave is 52/273 or 0.2 µs approximately, giving the highest fundamental video frequency of 5 MHz. Thus

$$f_{max} = \frac{\text{active lines} \times \text{Kell factor} \times \text{aspect ratio}}{2 \times \text{line forward scan period}} \qquad ...(30.4)$$

$$f_{max} = \frac{N_a \times K \times a}{2 \times t_{fH}}$$

where t_{fH} is the horizontal line forward scan period.

The fundamental law of communications, the *interchangeability of time and bandwidth* states that "if total number of units of information is to be transmitted over a channel in a given time by specified methods, a specific minimum bandwidth is required". If the time available for transmission is reduced by a factor, or if the total information to be transmitted is increased by the same factor, the required bandwidth is increased by this factor. Suppose for example, the picture detail in both horizontal and vertical directions is to be doubled without changing the frame rate. This represents an increase in total picture elements by a factor of 4.

INTERLACING OF SCANNING LINES

The *vertical resolution* of a system depends on the total number of active scanning lines, whereas the *critical flicker frequency* (CFF) is the lowest possible apparent picture repetition rate which may be used without visible flicker. If during one vertical scanning field *alternate* scanning *(odd)* lines are formed and during the second scanning field the *remaining (even)* lines are formed, at the end of the one frame period all the lines are formed. This process called *interlaced scanning* gives a vertical scanning frequency of twice the true picture-repetition, or frame frequency. The situation is illustrated in Fig. 30.10 and further elaborated in Fig. 30.11.

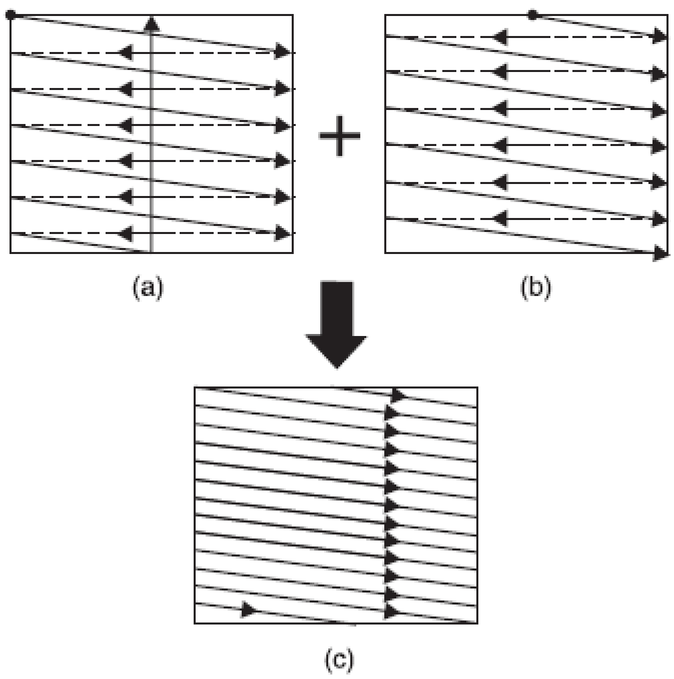


Fig. 30.10 One complete frame in interlaced scanning (a) Odd line field in interlaced scanning (b) Even line field in interlaced scanning (c) Odd-line field in (a) and even line field in (b) interlaced to form a complete frame

The upper frequency requirement of the video system is based on the true picture-repetition frequency, whereas the large area flicker effects are based on twice this frequency, or the vertical scanning frequency. The principle of *interlacing* the scanning lines, therefore, by cutting the complete picture repetition rate in half, *allows a system to maintain the same resolution and flicker characteristics as a noninterlaced system with a bandwidth requirement only half that of the noninterlaced system.*

Further improvement might result from more complicated interlacing systems. The ratio of field to frame frequency cannot be too great, however, since flicker would begin to show up for small picture areas which would not be covered by all the successive fields making up a complete frame.

The television picture is different from the picture projected in a movie theatre in that it consists of many lines. Another difference is the fact that a projected movie frame appears as a complete picture during each instant it is flashed on the screen. *The television picture is traced by a spot of light.* When the spot traces a television picture, it does so in the following manner;

1. The spot traces line 1 of the picture
2. The spot traces line 3 of the picture
3. The spot traces line 5 of the picture……and so on.

The spot traces all the *odd lines* of the picture. When the spot reaches the bottom of the picture, it has traced all the odd lines i.e. a total of 312.5 lines in the 625 line system. This is called the *odd-line field.* Fig. 30.11 (a).

After the spot has traced all the odd lines, the screen is blanked and the spot is returned quickly to the top of the screen once again. The time during which the spot travels back to the top of the screen is called the *vertical retrace period.* The timing of the this period is such that the spot reaches the *top centre* of the screen when it again starts tracing lines in the picture as shown in Fig. 30.11 (b). On this second tracing, the spot traces,

1. line 2 of the picture
2. line 4 of the picture
3. line 6 of the picture… and so on.

When the spot completes the tracing of all *even lines i.e.* a total of 312.5 lines in the 625 line system and is at the bottom of the picture, it has traced the *even-line field*, depicted on Fig. 30.11 (c).

The screen is then blanked once again, and the spot is returned to the *top-left-hand corner* of the screen, Fig. 30.11(d) and the tracing sequence repeats. *The method of scanning, where the odd lines and even lines are alternately traced, is called interlaced scanning.*

After both the odd and-even-fields have been scanned, the spot has given us *one frame* of the television picture. The word frame is borrowed from motion pictures, to indicate one complete picture that appears in a split second. *In television, the odd-line field and the even-line field make one frame.*

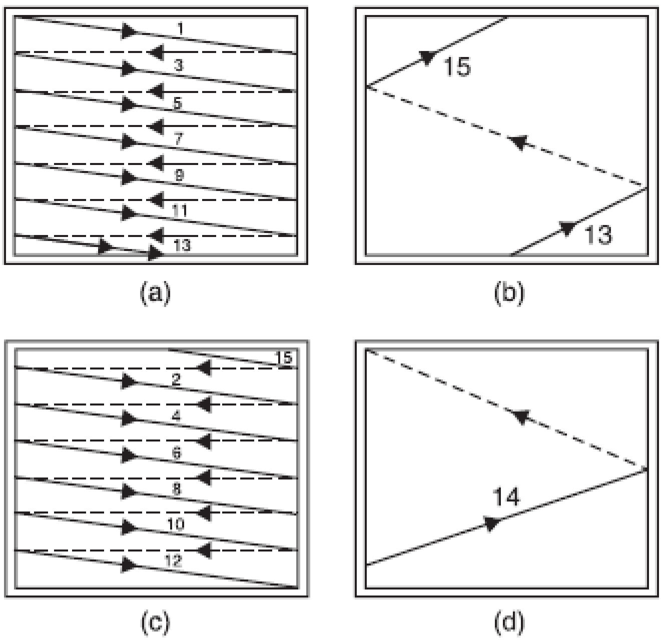


Fig. 30.11 Interlaced scanning; (a) odd line field (b) transition period (c) even line field and (d) transition period

TEST CARD

All TV stations transmit a *tuning signal* for some time before the start of the program proper, *to facilitate the tuning in of receivers*. The signal takes the form of a *Test Card* which, besides other geometrical figures, such as checker board pattern, always includes one or more circles and both horizontal and vertical patterns of lines. The circles and other geometrical figures aid in the correction, where necessary, of frame linearity and of aspect ratio. The vertical pattern of lines serves for the fine tuning of channel selector, and the horizontal one for the accurate adjustment of the spacing. Fig. 30.12 shows a test card offering a particularly wide range of possibilities.

You can receive the correct aspect ratio by adjusting for a pattern which resembles, Fig. 30.12 as closely as possible. The wedges in the test pattern should be as close to equal length as possible; the circle should be as round as possible.

THE VIDEO SIGNAL

The voltage equivalent of any horizontal line that is scanned *begins* where the electron beam starts scanning the image horizontally, and *ends* where the scanning finishes and the horizontal retrace blanking begins. If we assume that the scanning action continues line after line, as happens in television transmitters, we have delivered to the video preamplifier in the camera, a train of picture voltages with intervals of blanking in between. Although the lines scanned are positioned *one below the other*, the voltage representation of the these lines is a *train* of voltage fluctuations, *one following the other*, along a time base. A blanking voltage is applied to the camera tube to extinguish the beam during horizontal retrace, although the signal output from the camera is zero during the blanking interval. In other words, the blanking signal fed to the electron gun of the camera tube does not appear in the camera tube video output. *The voltage representation of every line scanned is separated from the next by the blanking interval*, Fig. 30.13.

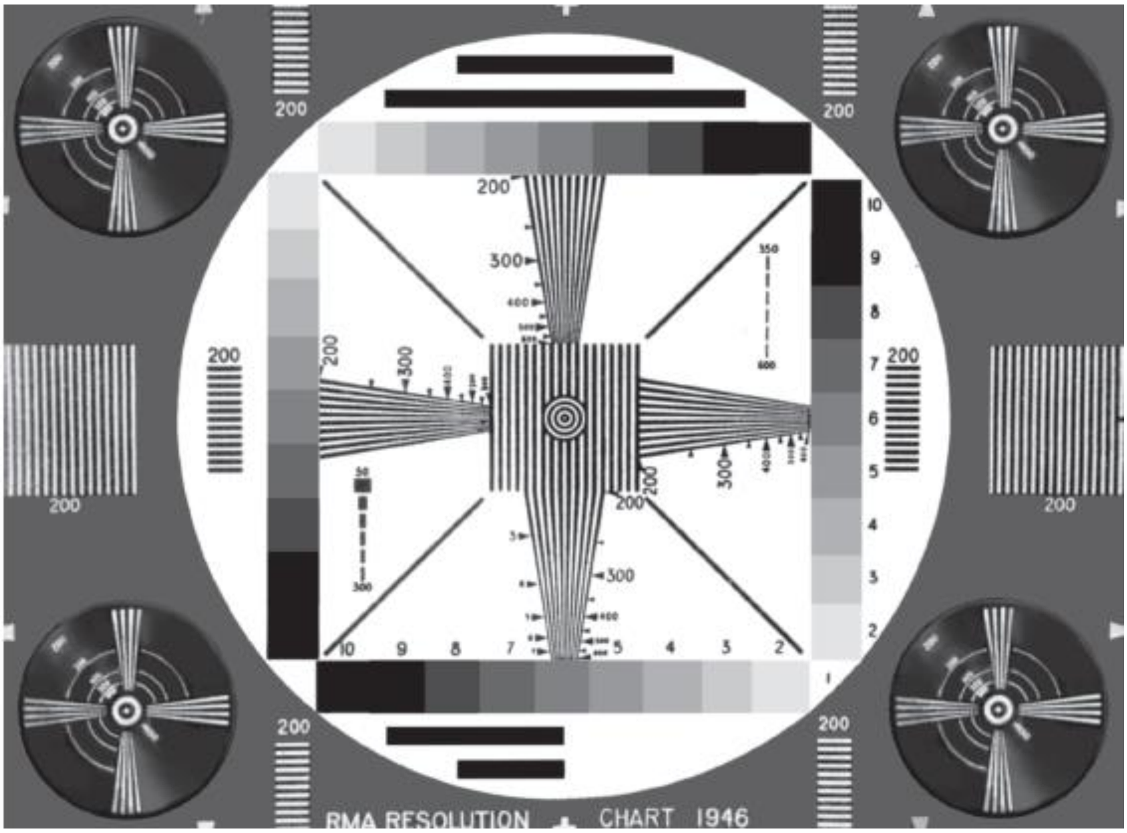


Fig. 30.12 Test Card

The video output from the iconoscope increases in amplitude as the image changes from black to white and white in the image produces stronger video signal. This is called *positive video polarity*. In negative video polarity the conditions of operation call for white in the image to produce the theoretically zero signal and black in the image to produce theoretically maximum video signal. The condition of *negative video polarity* can be met with by applying the iconoscope camera output signal to a negative clamping circuit. This is shown in Fig. 30.14. The whole of the signal then appears below the zero-voltage baseline and is entirely in the zone of negative polarity voltage.

When the fixed negative bias is added to the positive-polarity camera signal, the result is a voltage variation wherein white in the image becomes the *least negative* voltage and black in the image becomes the *most negative* voltage.

Before broadcast, the video signal is passed through a number of amplifying stages employing transistors. *Each amplifying stage inverts the direction of video signal.* It is quite obvious, that a positive polarity video signal can also be converted into a negative-polarity video signal, simply by employing an odd number of amplifying stages.

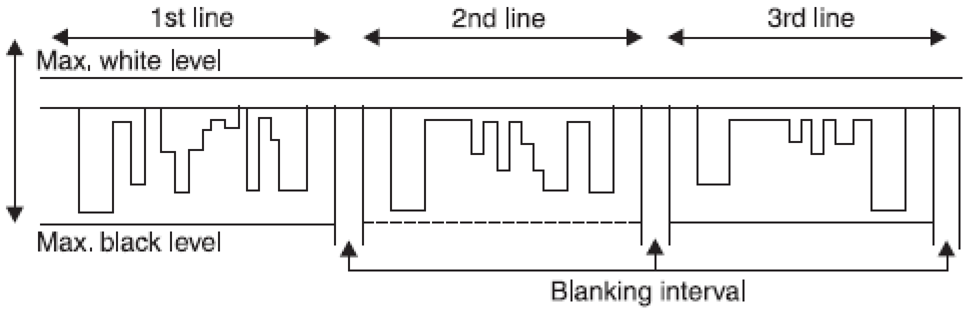


Fig. 30.13 Blanking interval (positive video polarity)

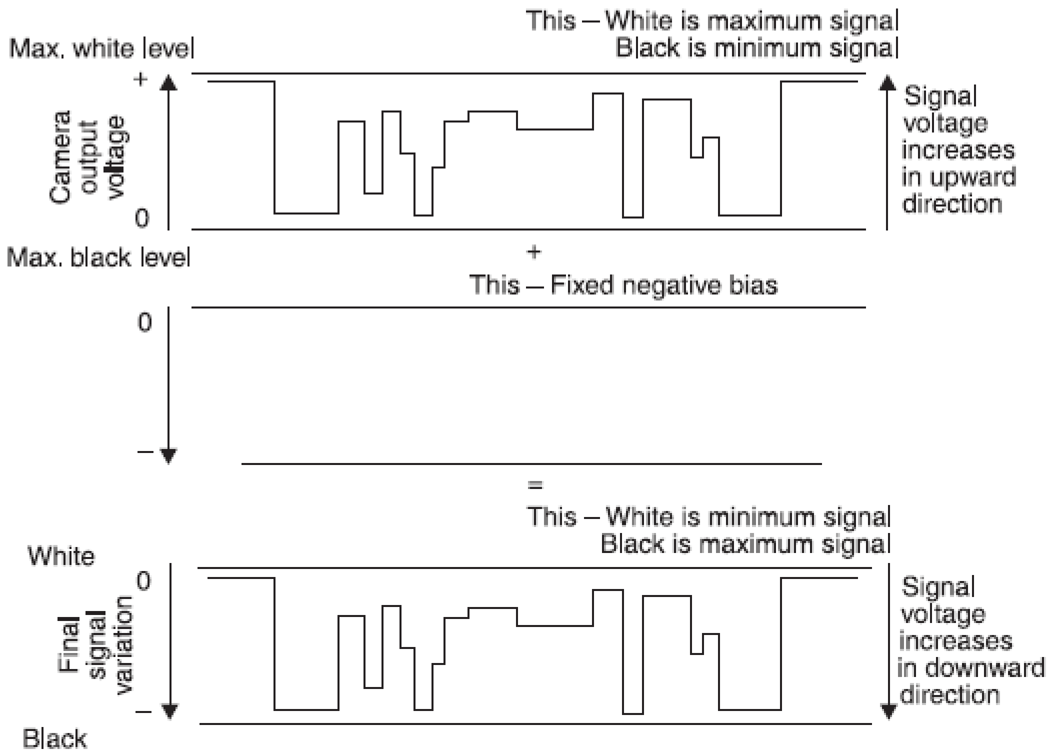


Fig. 30.14 Negative video polarity

CONTROL PULSES

TV receivers work precisely to give you a rock steady, crisp and clear picture, accompanied by scintillating melodious sound. You might think that all this calls for a precise, sophisticated and expensive, TV receiver. *No doubt, precision is essential in television but it is in the transmitting equipment and not in the receiving equipment.* Whether this precision is in the transmitting equipment, or in the receiving equipment, it is immaterial, because anyway it serves the same purpose. To maintain this precision in the transmitting equipment is, of course, convenient, practical, and economical. Where this precision maintained in the receiving equipment, TV receivers would have become complicated, costly, and beyond the repair capability of each and every technician.

In addition to the normal video and sound, the transmitted signal also contains the following *control signals:*

(1) line blanking pulses; (2) frame blanking pulses; (3) line synchronising (sync) pulses; (4) frame synchronising (sync) pulses: (5) pre-equalising pulses; and (6) post-equalising pulses.

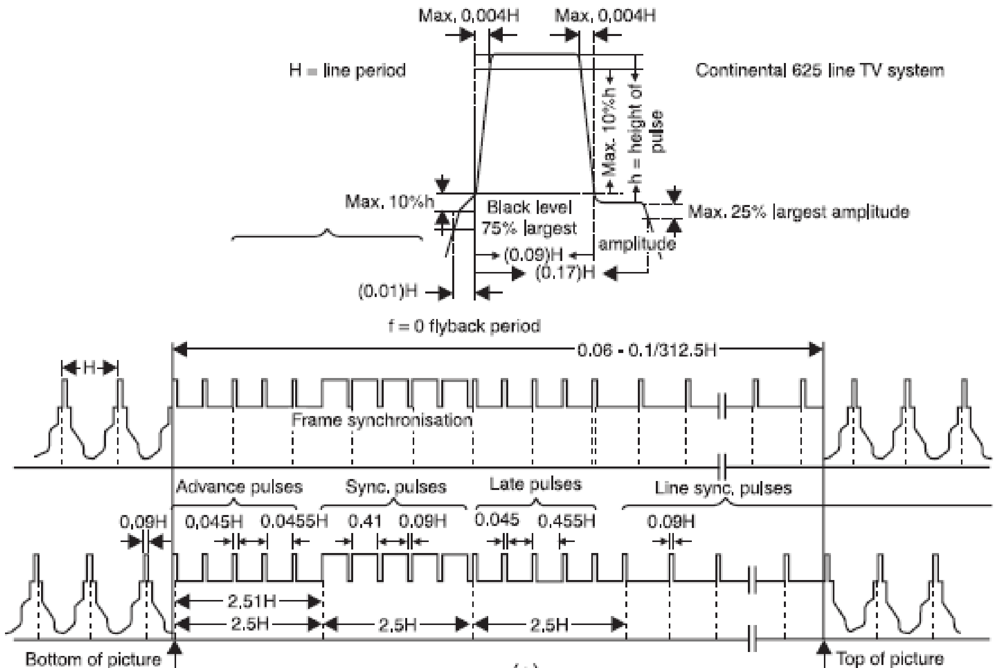
The video, sound, and control signals combine to make up the composite video signal. Within the TV receiver, the circuits are slaved to this signal. For example, the frame oscillator, if left free, would drift in frequency, resulting in vertical instability. Similarly, the line oscillator, if left free, would also drift in frequency, resulting in horizontal instability. *The composite video signal locks in the frequency and phase of the frame and line oscillators.*

COMPOSITE VIDEO SIGNAL

The simple video signal we started with has turned into a video signal which is really complex. The variations of current strength conveying the picture content are here interwoven with a dense system of pulses, Fig. 30.15(a), all working towards a common aim. *If this train of pulses has to be transmitted, a method that suggests itself is to modulate the amplitude of the carrier wave in the rhythm of the pulses.*

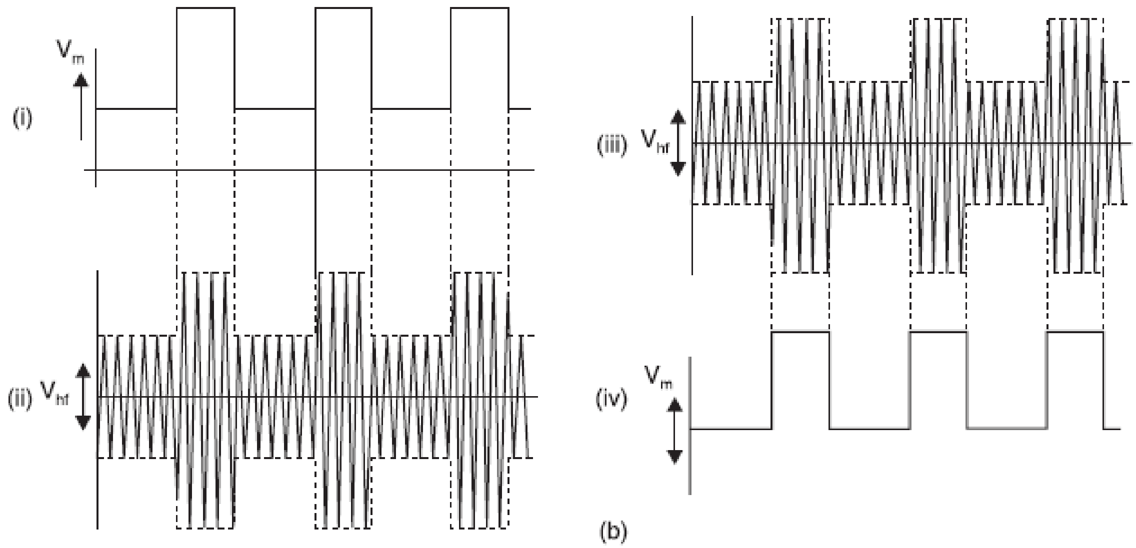
There will have to be an amplitude detector in the receiver to recover this pulse train from the modulated carrier wave. The detector consists of a simple network that transforms high-frequency electrical waves into a direct voltage. The magnitude of the direct voltage produced always corresponds exactly to the strength of the high-frequency wave, and thus it will give an accurate representation of the changes in the amplitude of the wave. The output voltage of the detector will thus have the same pulse like character as the voltage modulating the carrier at the transmitter, Fig. 30.15(b). *This method of modulation, whereby the amplitude of the carrier wave is made to vary between a maximum and a minimum is called amplitude modulation (AM).*

An important characteristic of amplitude modulation is the *degree of modulation;* it is the difference between maximum and minimum amplitudes of the carrier expressed as percentage of its average amplitude. Sinusoidal modulation with 50%, 80% and 100% depth of modulation is illustrated in Fig. 30.15(c). Full modulation of a carrier wave, to a depth of 100%, is obviously the most efficient, but it is avoided in radio and TV transmitters because it makes accurate demodulation in the receiver a very difficult matter. The picture carrier as modulated with the video signal (negative modulation) is shown in Fig. 30.15(d).



Max. 0.004H
Max. 0.004H
H = line period
Continental 625 line TV system
Max. 10%h
h = height of pulse
Max. 10%h
Black level 75% largest amplitude
Max. 25% largest amplitude
(0.09)H
(0.17)H
(0.01)H
f = 0 flyback period
0.06 - 0.1/312.5H
H
Frame synchronisation
Advance pulses
Sync. pulses
Late pulses
Line sync. pulses
0.09H
0.045H
0.0455H
0.41
0.09H
0.045
0.455H
0.09H
2.51H
2.5H
2.5H
2.5H
Bottom of picture
Top of picture

(a)



(i)
V_m
(ii)
V_{hf}
(iii)
V_{hf}
(iv)
V_m

(b)

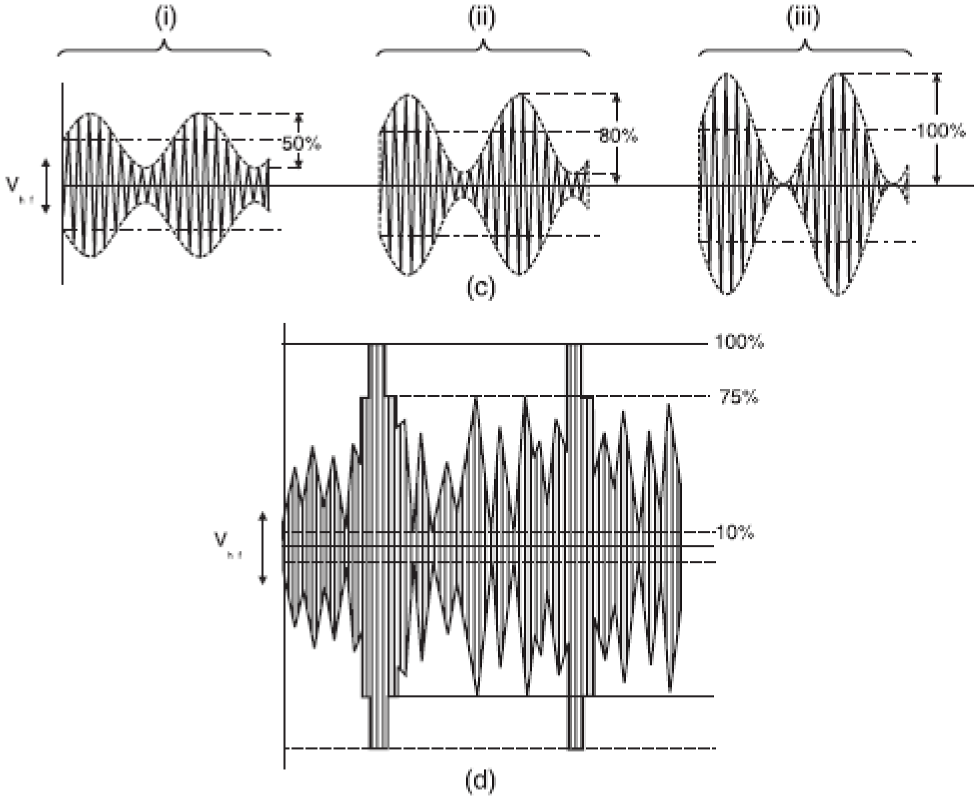


Fig. 30.15 Composite video signal (a) The train of pulses in a composite video signal. (b) (i) The pulse train as modulation voltage ; (ii) the amplitude modulated carrier wave. (iii) carrier wave with pulse-shaped modulation; (iv) pulse train as restored by the demodulation process in the receiver. (c) Depth of modulation in AM; a carrier wave modulated to depths of (i) 50% (ii) 80% and (iii) 100%. (d) The vision carrier as modulated with the video signal (negative modulation). The levels are indicated as percentages of the maximum amplitude of the carrier; white level 10%, black level 75%, “blacker-than-black” region 75-100%.

THE HIGHEST VIDEO FREQUENCY

For the picture to be reproduced with sufficient sharpness and wealth of detail, its area must be split up into about 400,000 elements. At the rate of 25 pictures a second, a total of 10,000,000 of these elements will have to be scanned per second. Let us assume an extreme case in which all the elements are alternately bright and dark. The picture signal—the voltage derived from the camera tube—would then be a toothed wave, with five million positive teeth and five million negative gaps, part of which is shown in Fig. 30.16.

Let us further replace this toothed wave with a sine wave (dotted line). This sinusoidal oscillation represents a frequency of 5 MHz. An alternating voltage, with this frequency, is the most rapidly changing one that can arise from the scanning of a picture. This is the highest video frequency of the system and plays a dominant role in the whole chain of transmission, from the scene being televised in the studio to the scene being viewed on the picture tube screen. *Obviously, the highest video frequency depends on the number of*

THE LOWEST CARRIER FREQUENCY

The carrier wave itself must have a frequency far higher than the highest video frequency, otherwise, the shape of the modulation would not be clearly reflected in the modulated carrier. This will be clear from Fig. 30.17.

With a carrier frequency of at least five times as great as the modulation frequency, the modulation shape finds clear expression in the modulated carrier, Fig. 30.17 (a). On the other hand, when carrier frequency is only twice modulation frequency, it is scarcely possible to recognise the shape of the modulation, as shown in Fig. 30.17 (b).

The carrier frequency chosen is, in practice, at least five times as great as the modulation frequency. For television, carrier waves must be used with a frequency higher than about 30 MHz. Colour television transmission only becomes possible in the VHF and higher ranges—that is, on wavelengths under 10 metres. In fact, *the lowest frequency officially allocated to television by international agreement is over 40 MHz.*

SIDE BAND FREQUENCIES

The basic form of any process, that repeats itself periodically, is a pure sine wave of constant amplitude and frequency. Processes of a complicated nature, with amplitude or frequency changing in a haphazard manner, are composed of several elementary sinusoidal oscillations. If a 20 kHz wave is amplitude modulated with a 4 kHz wave, there exist in reality three oscillations having constant amplitude: a stronger oscillation of 20 kHz, and two equally weak oscillations of (20 – 4) = 16 kHz, and (20 + 4) = 24 kHz. The three of them appear one above the other. Adding together the amplitudes of the three waves, *point by point*, results in the amplitude modulated wave, Fig. 30.18.

Amplitude modulation does not signify the alteration of the carrier wave amplitude, but what happens is that two new waves are formed, arising *symmetrically* on either side of the carrier. The two new waves are generated automatically in the modulation stage of the transmitter. *The new waves arising in this way, are called sideband waves or are referred to collectively as the upper and lower sidebands of the carrier wave.*

What happens when the carrier is modulated by some other voltage shape, for example, a video signal which, of course, is not sinusoidal. *Completely irregular, as the video information is, and dependent only on differences of brightness between individual picture elements, the video signal is, in reality, made up of a great number of pure sinusoidal oscillations, having various amplitudes and frequencies. This mixture of frequencies consists of two distinct groups of waves.* The *first group* includes all waves with a frequency that is a *multiple of the line frequency.* In the 625 line system, the line frequency is 25 × 625 = 15,625 Hz, and this group, therefore, includes the following frequencies 15,625 Hz; 2 × 15,625 = 31,250 Hz; 3 × 15,625 = 46,875 Hz; 4 × 15,625 = 62,500 Hz and so on. The wave of frequency 15,625 Hz is called the *fundamental.* The waves having twice, three times, and four times the fundamental frequency are called the second, third and fourth *harmonics of the fundamental.*

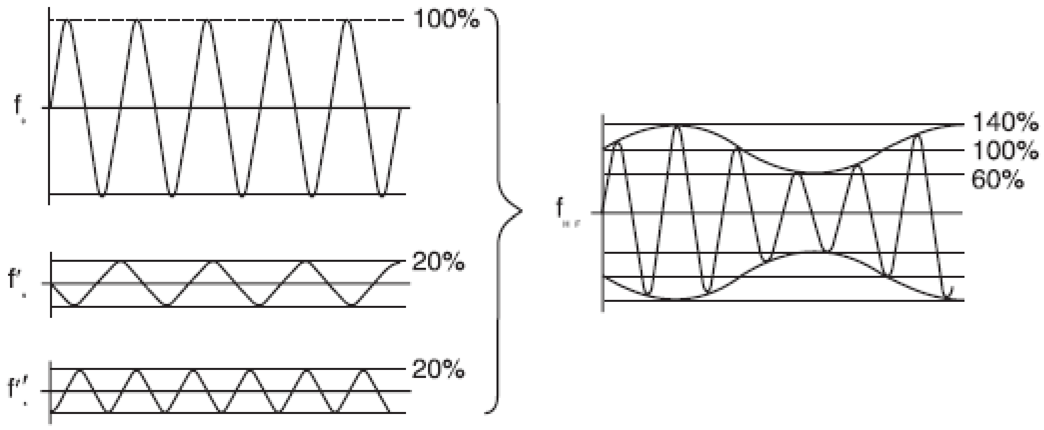


Fig. 30.18 Side waves arising when a 20 kHz carrier is amplitude modulated with a wave of frequency 4 kHz

The *second group* of waves, contained in the video signal covers the *picture frequency of 25 Hz, and its harmonics*, i.e, 50,75,100 Hz and so on. In the composite video signal, the two groups of waves are so combined that an upper and lower sideband each consisting of a complete group of waves, made up of the picture frequency and its harmonics, is *arranged symmetrically* around each harmonic of the line frequency i.e. at 15,625; 31,250; 46,875 Hz etc. This is illustrated in Fig. 30.19.

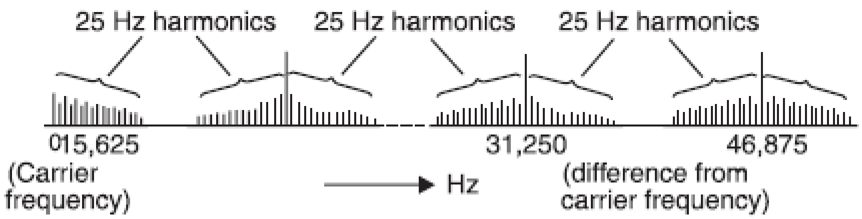


Fig. 30.19 The two groups of sideband frequencies of amplitude modulated video signal

The waves of the first group are the predominant ones, while the sidebands of the 25 Hz group play only a subordinate part and quickly tail off in amplitude. This frequency spectrum, as it is called, exhibits the enormous number of pure sinusoidal oscillations, that go to make up the composite video signal, resulting from the scanning of a picture. *We need take into account only frequencies upto 5 MHz.* The finest picture detail is carried by still higher frequencies, but these are beyond the power of resolution possessed by the human eye.

A whole band of frequencies is produced in the modulation stage of a transmitter. The whole of this band has to be *radiated* by the transmitting antenna, *taken up* by the receiving antenna, and *amplified* in the receiver, in order that it can be finally changed back to its original form, the video signal in the demodulation stage.

This emphasizes the necessity for all the links in the chain of transmission to have sufficient bandwidth. This means that the transmitting and receiving antennas must respectively radiate and take up all the waves from the lowest to the highest sideband frequency, with the same efficiency; it also means that the receiver must amplify all these frequencies to the same extent. If this is not so, and some of the waves are somehow handicapped during their journey, then the picture signal recovered in the demodulation stage will have *a form different from the original.* As an example, a weakening of the lowest or highest sideband frequencies will result in a loss of picture detail.

FREQUENCY MODULATED SOUND CARRIER

The sound waves picked up by the microphones, are translated into electrical voltages by these devices, Fig. 30.20(a), and fed into the control room through individual audio cables. The sound modulated current is amplified to many times its original strength and then conveyed to the transmitter for impressing on the carrier wave. The sound signal could be impressed on the carrier by amplitude modulation (AM). *In the 625 line system, the sound carrier is frequency modulated.* The difference between amplitude modulation (AM) and frequency modulation (FM) is shown in Fig. 30.20.

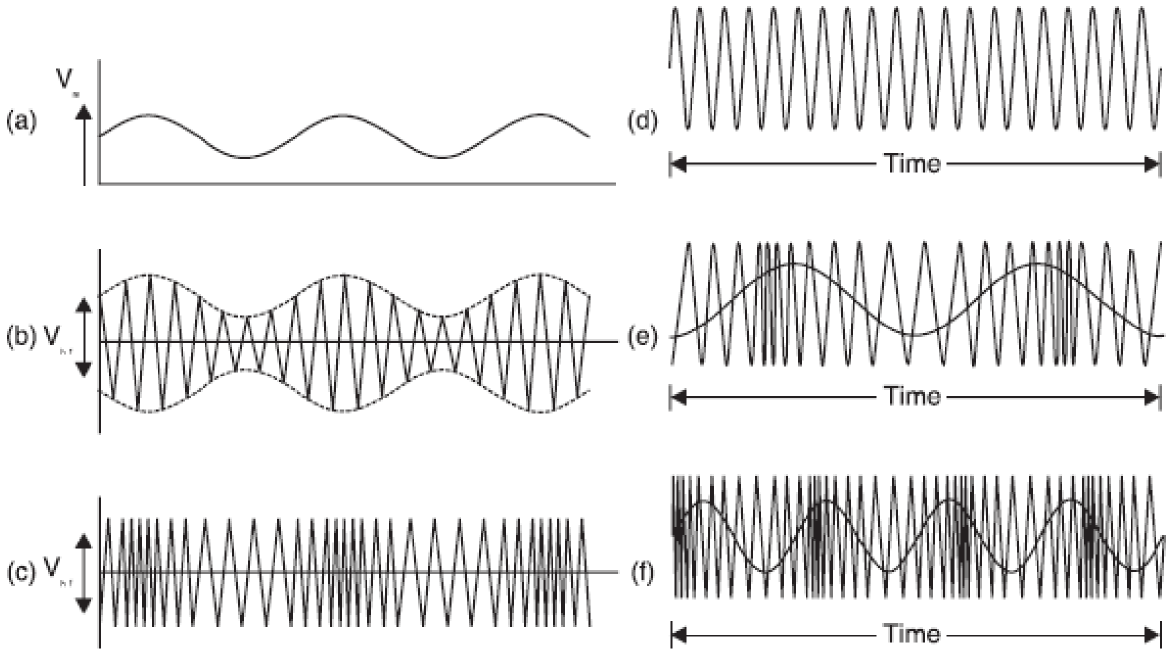


Fig. 30.20 Amplitude and frequency modulation (a) V_m—A sinusoidal modulation voltage. (b) A carrier wave, V_h, amplitude modulated by V_m (c) A carrier wave, V_{hf}, frequency modulated by V_m (d) Resting frequency (e) Deviation produced by a whisper (f) Deviation produced by a bang

In frequency modulation the amplitude of the carrier remains unchanged, while its frequency is altered in the rhythm of the modulating wave. In the absence of any modulation, the radiated sound carrier has the constant or resting frequency, Fig. 30.20(d). *The change in carrier frequency during frequency modulation is known as frequency deviation.*

When the frequency *increases above* the resting frequency, it is known as *positive deviation;* when the frequency *decreases below* the resting frequency, it is called *negative deviation.* The change in frequency is called *swing.* This is illustrated in Fig. 30.20(e) and (f).

In practice it is necessary to lay down a *maximum frequency swing* for all FM systems just as it is necessary to lay down a *maximum percentage modulation* for all AM systems. This serves as a basis for the design of FM detectors in receivers. The detectors must work in such a way that, with maximum swing, it delivers the maximum permissible voltage to the loudspeaker or the output stage, whatever the case may be. *The maximum deviation laid down for television sound is + 50 kHz.*

In general, frequency modulation in the VHF range is an easier matter than the accurate and satisfactory amplitude modulation of the same high frequencies. The shorter the wavelength used, the stronger the undersirable effect that the modulating signal has on the frequency of the carrier; in the end it becomes simpler to incur the slight expense of turning these almost unavoidable fluctuations of the carrier frequency into proper frequency modulation, rather than spend a great deal more on suppressing them. For this reason, *transmitters in the centimetre wave range work almost exclusively on frequency modulation.*

The sidebands arising in FM, constitute a very important problem. In FM, an infinite number of sideband frequencies arise *symmetrically* around the carrier frequency even when it is only a single sine wave that is modulating the carrier. The *sidebands* are separated by a frequency which is always equal to the modulating frequency; at first, with increasing distance from the carrier, their amplitudes vary irregularly, then they fall off regularly and rapidly, as shown in Fig. 30.21 (a). The modulating voltage consists of the sum of several sine waves. Each one of these sine waves will generate a sideband spectrum of its own. The complete frequency spectrum will be extraordinarily extensive and complex. Theoretically, all of these sidebands must reach the detector stage in the receiver, in order that they may be translated into a faithful reproduction of the original sound.

The question now arises, *how many* sidebands are essential for a sufficiently faithful reproduction of the original sound, and how many of them will have to reach the detector. It depends on the ratio between the modulating frequency and the frequency swing. As a thumb rule, one may say that *at least first five to seven paired waves of each sideband must be transmitted.* Because the frequency separating sideband waves is always equal to the modulation frequency, each group of seven sidebands will extend over a frequency band equal to seven times the modulation frequency, the two bands lying on either side of the carrier. If the modulation frequency has the highest value that has to be transmitted, i.e. 15 kHz (the highest audio frequency), the distance of both the *highest* and the *lowest* sideband frequencies from the carrier will be (7 × 15) = 105 kHz, Fig. 30.21 (b).

We conclude that *the bandwidth of every link in the chain of FM transmission must be at least 200 kHz.* This is a great disadvantage in comparison with any AM system, in which only one upper and one lower sideband wave has to be transmitted and which, for conveying the maximum audio frequency of 15 kHz; only requires a bandwidth of (2 × 15) = 30 kHz. In actual practice, the bandwidth of AM transmitters is often limited to 9 kHz, making the highest audio frequency transmissible 4.5 kHz.

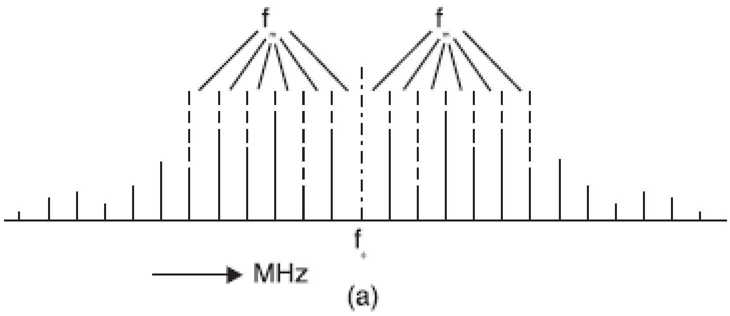


Fig. 30.21 The sidebands of an FM carrier (a) f_o = Carrier wave (central frequency) ; f_m = separation between the frequencies of individual side waves. The separation is always equal to the modulation frequency.

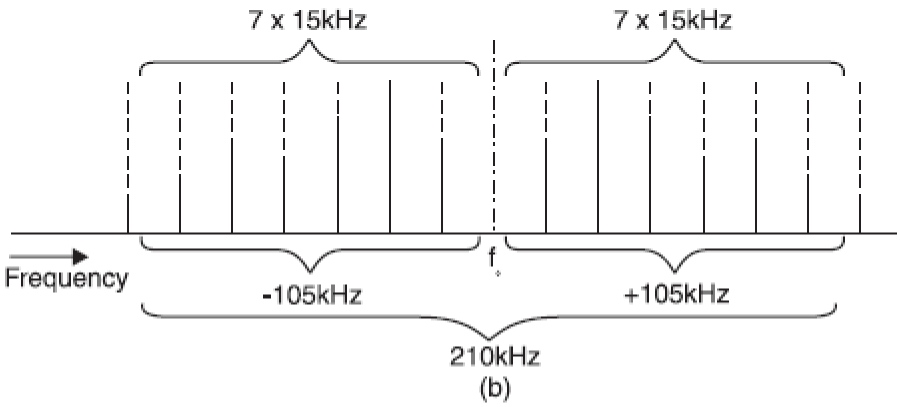


Fig. 30.21 (b) For sound frequencies upto 15 kHz to be reproduced without distortion it is necessary to transmit sidebands extending to about 105 kHz on either side of the carrier frequency.

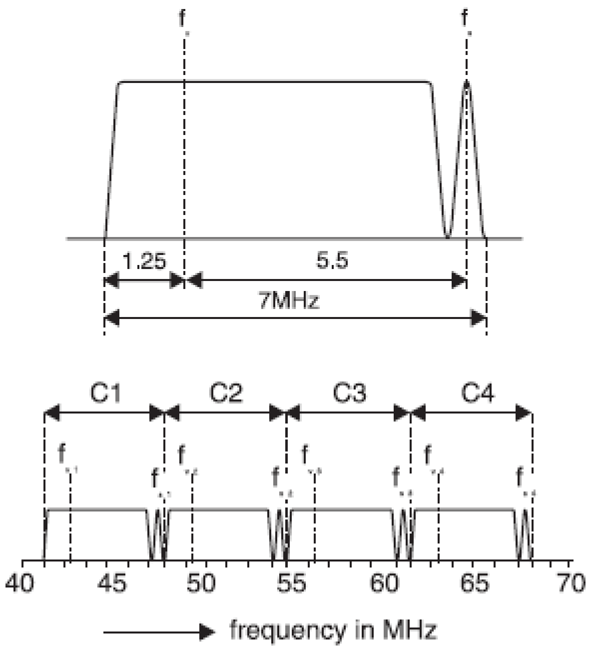


Fig. 30.22 A television channel as showing the frequency characteristics of the vision and sound transmitters. f_v = vision carrier; f_s = sound carrier

In consequence of the great bandwidth they require, FM systems can only be employed in very high frequency (VHF) ranges, in which a broad frequency band can be allocated to each transmitter without danger of its breaking into the sideband of some other transmitter. It is impossible to do this in the long, medium or short wave bands; it may become possible in the VHF range, that is, with wavelengths under 10 metres and frequencies over 30 MHz.

A sound signal is broadcast at the *same time* as the picture signal; silent television has never existed. The information in the sound signal is quite separate and different from that of the picture signal. The simplest course is to convey it by means of a second carrier wave, and *have it in the same frequency range as the picture carrier; and as close to it as possible.* The reasons for doing so are to make it possible for a single receiving aerial to take up both picture and sound systems efficiently and to enable both signals to be amplified together, at least in the earlier stages of the receiver. *The sound carrier frequency is, therefore, given a frequency such that it lies just far enough above the highest picture sideband frequency, to ensure that the much narrower sidebands of the sound carrier do not interfere with the picture sideband.* A separation of about 1 MHz is sufficient to avoid such interference. A 625 line television channel showing the pass bands of the vision and sound transmitters is shown in Fig. 30.22. The term *sound channel* is frequently used for the narrow frequency band taken up by the sound transmitter within the television channel.

MONOCHROME TV CAMERA

Simplified diagram of a typical monochrome television camera is given in Fig. 30.23. An *optical system* focuses light reflected from the scene onto the face plate of the camera tube. A*photoelectric process* transforms the light image into a virtual electronic replica in which *each picture element is represented by a voltage.* A *scanning beam* in the pick-up tube next converts the picture, element by element, into electrical impulses. At the output an *electrical sequence* developes that represents the original scene. The output of the camera tube is then amplified to provide the video signal for the transmitter. A sample of the video signal is also provided for observation in a CRT viewfinder mounted on the camera housing.

Electronic circuits that provide the necessary control synchronisation and power supply voltages operate the TV camera tube. A *deflection system* is included in the TV camera to control the movement of the camera tube scanning beam. Many TV cameras receive synchronising pulses from a *studio control unit.* This unit also provides the sync pulses that synchronises the receiver with the camera.

Some TV cameras however generate their own *control signals.* In turn, they provide output pulses to *synchronise* the control unit. Manual controls are also provided at the rear of the camera for setting the optical lens and for zooming. Because of the complex electronic circuits and controls, early TV cameras were rather large and awkward to handle. Recent innovations have revolutionised the cameras' construction.

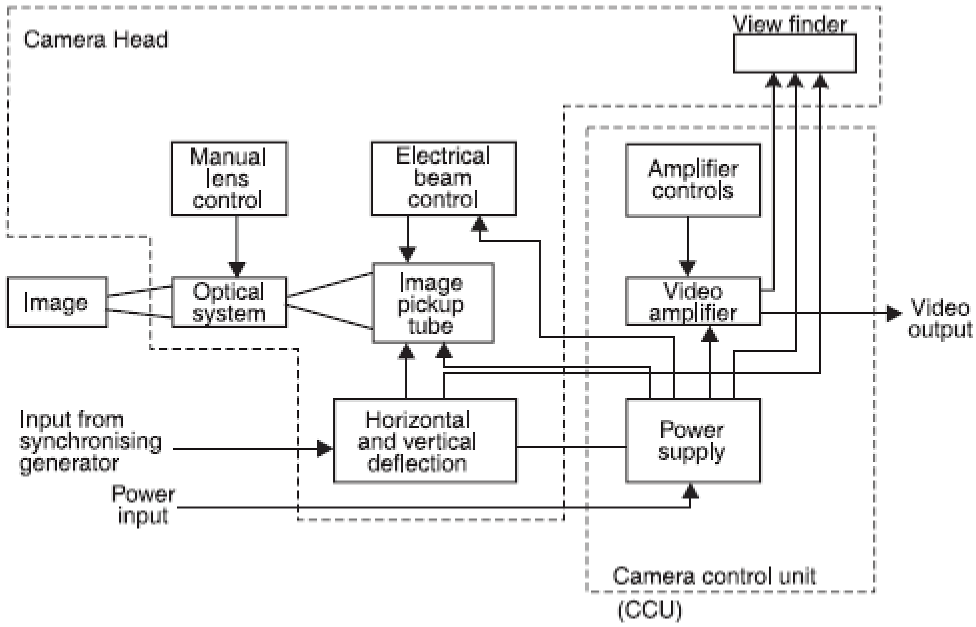


Fig. 30.23 Simplified diagram of a typical monochrome television camera

CAMERA TUBE CHARACTERISTICS

One important characteristic of all camera tubes is their *light transfer characteristic.* This is *the ratio of the face plate illumination in foot candles to the output signal current in nanoamperes (nA).* This characteristic may be considered as a measure of the efficiency of a camera tube. Typical values of output current range from 200 nA to 400 nA.

The term *gamma*, in a television system, is applied to camera tubes and to picture tubes. *Gamma is a number which expresses the compression or expansion of original light values.* Such variations, if present, are inherent in the operation of the camera tube or picture tube.

With *camera tubes* the gamma value is generally 1. This value represents a *linear characteristic* that does not change the light values from the original scene when they are translated into electronic impulses, see Fig. 30.24(a). However, the situation is different in the case of *picture tubes* which have gamma of approximately 3. The number varies slightly for different types of picture tubes and also for different manufacturers. For picture tubes it is desirable to provide *improved contrast.* Emphasising the bright values to a greater degree than the darker values accomplishes this. Fig. 30.24(b) is a typical picture tube gamma characteristic. Note that *the bright portions of the signal operate on the steepest portion of the gamma curve.* Conversely, *the darker signal portions operate on a lesser slope.* Thus it can be seen that the bright picture portions will be emphasised to a greater degree than the darker picture portions.

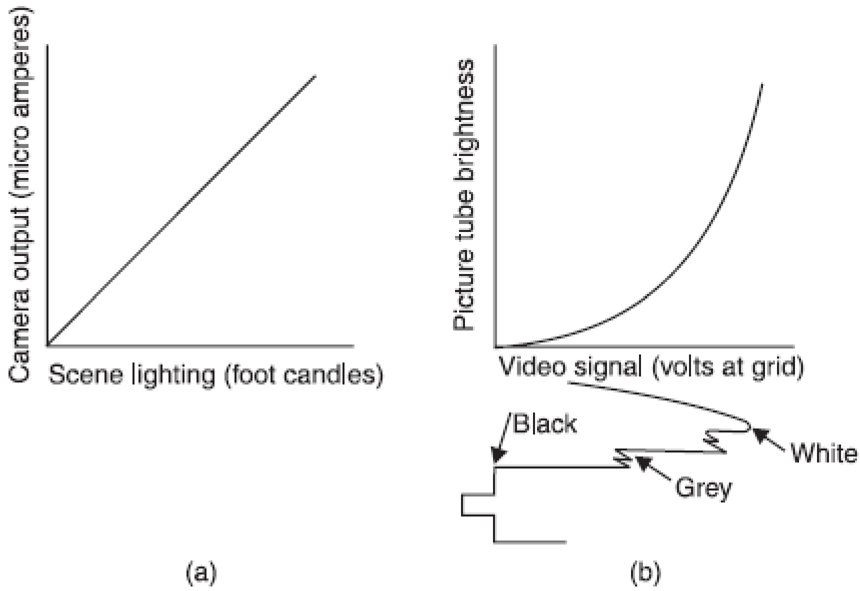


Fig. 30.24 Gamma characteristic of a camera tube

For proper colour reproduction the *spectral response* of a camera tube is an important parameter. As nearly as possible the tube should have the same spectral response as the human eye. This is necessary to render colours in their *proper tones.* It is also important in reproducing black and white pictures, thereby producing the *proper grey scale.* Tubes designed to operate in a colour camera have a greater response to each of the primary colours. Today, spectral response distribution has made possible the manufacture of camera tubes that are sensitive to the infrared, the ultraviolet and even the X-rays. But variations in spectral response have had little effect on the other operating characteristics of the tube.

If the photosensitive material in a camera tube was able to emit an electron for each photon of light focused upon the material, the *quantum efficiency* of the material would be 100 per cent. The formula for quantum efficiency is :

$$Q_{eff} = \frac{\text{electron}}{\text{photons}} \qquad \text{...(30.5)}$$

A quantum efficiency of 100% is almost impossible. However *quantum efficiency is a practical way to compare photosensitive surfaces in a camera tube.* In this comparison, photocurrent per lumen is measured using a standard light source. A *lumen* is the amount of light that produces an illumination of one foot candle over an area of one square foot. The source adopted for measurement is a tungsten-filament light operating at a colour temperature of 2870°K. Since the lumen is actually a measure of brightness stimulation to the human eye, *quantum efficiency is a convenient way to express the sensitivity of the image pick-up tube.*

Our eyes operate as a frequency-selective receiver that peaks in the yellow-green region at a wavelength of about 560 nanometres (560×10^{-9} metres) as shown in Fig. 30.25. Spectral response of vidicon type 7262A is illustrated in Fig. 30.26.

The term *lag* refers to the time lag during which the image on the camera tube decays to an innoticeable value. All camera tubes have a tendency to *retain* images for short periods after the image is removed. Some types do it more than others. *Lag on a television picture causes smear (comet tails) to appear following rapidly moving objects.*

Lag may be expressed as a percentage of the initial value of the signal current remaining 1/20s *after the illumination is removed.* Typical lag values for vidicon range from 1.5% to 5%. Plumbicons have lag values as low as 1.5%.

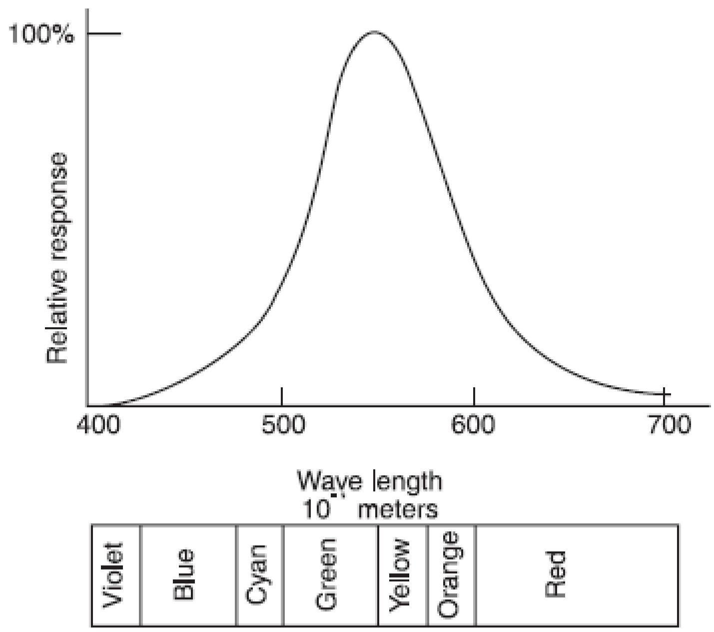


Fig. 30.25 Human spectral response

Dark current refers to the current that flows through the device even in total darkness. It is a form of semiconductor *leakage current.* Dark current forms the floor of the video signal or black level. Variations in *temperature*, therefore, alter the black level and electronic processing is required to maintain a *constant black level.* Automatic black level circuits clamp the black level to the setup level of the video signal.

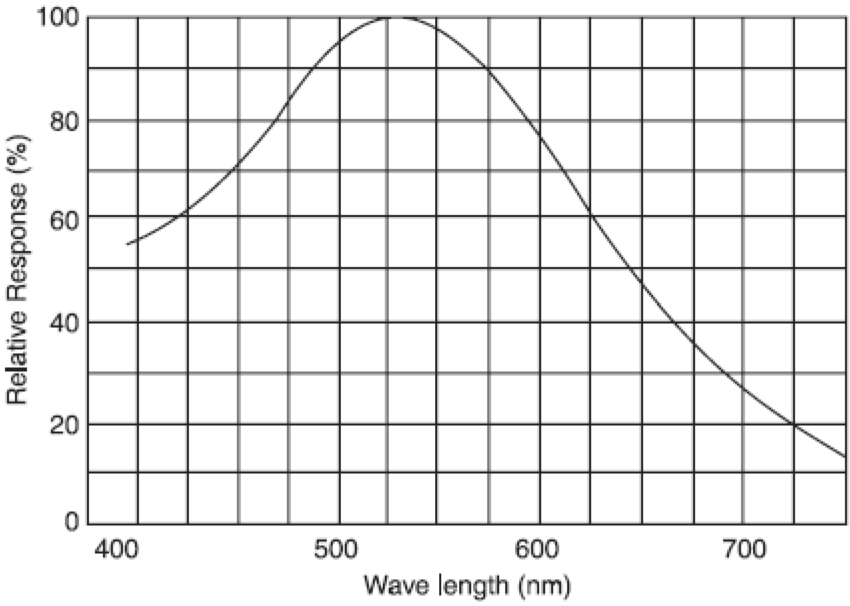


Fig. 30.26 Spectral response of vidicon type 7262A

Dark current also varies with applied *target voltage* in vidicons. One of the factors that lowers sensitivity is the *recombination* of electrons and holes where both cease to act as current carriers. By increasing the *voltage field gradient* in the target the likelihood of recombination decreases and more of the carriers liberated by light reach the target surfaces.

Sensitivity also varies with dark current and, therefore target voltage. Sensitivity rises for higher values of dark current. Dark current is adjusted to particular values by adjusting the target voltage. This means that sensitivity varies somewhat with the target voltage and a very simple automatic sensitivity control can be set up which resets target voltage as a function of output signal current.

Typical characteristics of vidicon camera tube are given in the Table 30.1.

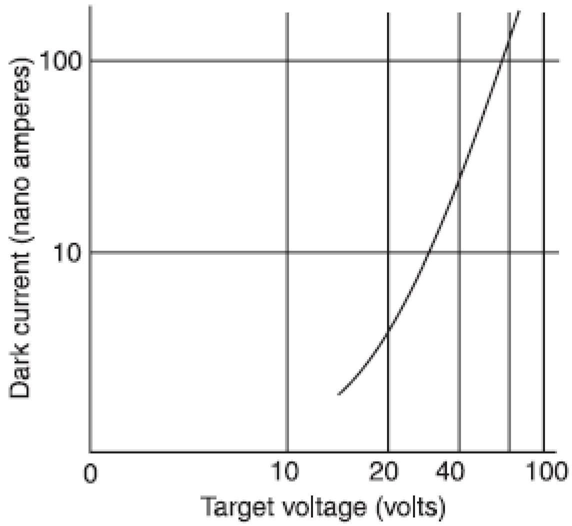


Fig. 30.27 Dark current versus target voltage

Table 30.1 Characteristics of Vidicon Camera Tube

Sensitivity	125 μA per lumen
Signal to noise ratio	45 dB
Image lag	Severe (about 20%)
Halo	No halo
Ghost	No Ghost
Dark Current	20 nA
Resolution	55% at 400 lines (5 MHz)
Spectral response	Close to eye's response
Infra-red sensitivity	Moderate
Gamma	0.7
Size	Small, compact (25 mm and 15 mm size target)
Applications	Outdoor recording, domestic recording

VIDICON CAMERA TUBE

Vidicon is a simple, compact TV camera tube that is widely used in education, medicine, industry, aerospace and oceanography. It is, perhaps, the most popular camera tube in television industry. During the past several years, much effort has been spent in developing new *photoconductive materials* for use in its internal construction. Today with these new materials, some vidicons can operate in exposure to direct sunlight or in near total darkness. Also, these tubes are available in diameters ranging from ½ to 4½ inches, and some of the larger ones even incorporate *multiplier sections* similar to those in the image orthicon. The main difference between the vidicon and the image orthicon is physical size and the photosensitive material that converts incident light rays into electrons.

The *image orthicon* depends on the principle of *photoemission* wherein electrons are emitted by a substance when it is exposed to light. The *vidicon*, on the other hand employs *photoconductivity*, that is, a substance is used for the target whose resistance shows a marked decrease when exposed to light.

In Fig. 30.28 the *target* consists of a transparent conducting film (the *signal electrode)* on the inner surface of the *face plate* and a thin photoconductive layer deposited on the film. *Each cross sectional element of the photoconductive layer is an insulator in the dark but becomes slightly conductive where it is illuminated.* Such an element acts like a leaky capacitor having one plate at the positive potential of the signal electrode and the other one floating. When the light from the scene being televised is focused onto the surface of the photoconductive layer, next to the face plate, each illuminated element conducts slightly the current depending on the amount of light reaching the element. This causes the potential of the opposite surface (towards the gun side) to rise towards the signal electrode potential. Hence, *there appears on the gun side of the entire layer surface a positive-potential replica of the scene composed of various element potential corresponding to the pattern of light which is focused onto the photoconductive layer*.

When the gun side of the photoconductive layer, with its positive potential replica, is *scanned* by the electron beam, electrons are deposited from the beam until the surface potential is reduced to that of the cathode in the gun. This action produces a change in the difference of potential between the two surfaces of the element being scanned. When the two surfaces of the element which form a *charged capacitor* are connected through the signal electrode circuit and a scanning beam, a current is produced which constitutes the *video signal.* The amount of current flow is proportional to the surface potential of the element being scanned and to the rate of the scan. The video signal current is then used to develop a *signal-output voltage* across the load resistor. The signal polarity is such that for highlights in the image, the input to the first video amplifier swings in the negative direction. In the interval between scans, whenever the photoconductive layer is exposed to light, migration of charge through the layer causes its surface potential to rise towards that of the signal plate. On the next scan, sufficient electrons are deposited by the beam to return the surface to the cathode potential.

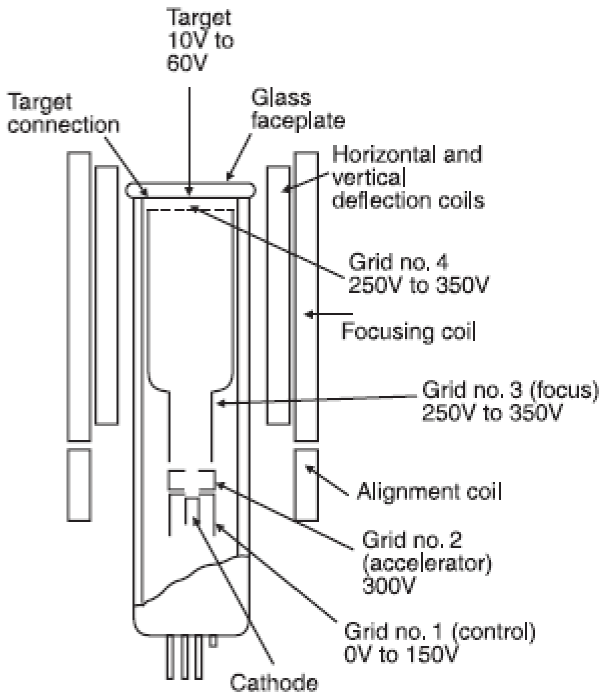


Fig. 30.28 Internal construction and external components of a vidicon camera tube. Voltages shown are typical.

The *electron gun* contains a cathode, a control gird (grid no. 1) and an accelerating grid (grid no. 2). The beam is focused on the surface of the photoconductive layer by the combined action of the uniform magnetic field of an external coil and the electrostatic field of grid no. 3. Grid no. 4 serves to provide a uniform decelerating field between itself and the photoconductive layer, so that the electron beam will tend to approach the layer in a direction *perpendicular to it*, a condition that is necessary for driving the surface to cathode potential. The beam electrons approach the layer at a *low velocity* because of the low operating voltage of the signal electrode. *Deflection* of the beam across the photoconductive substance is obtained by external coils placed within the focusing field.

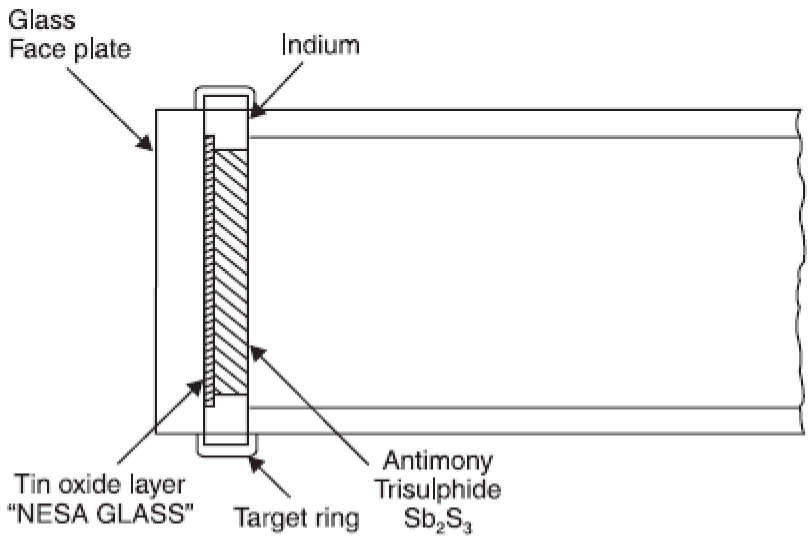


Fig. 30.29 Vidicon target assembly

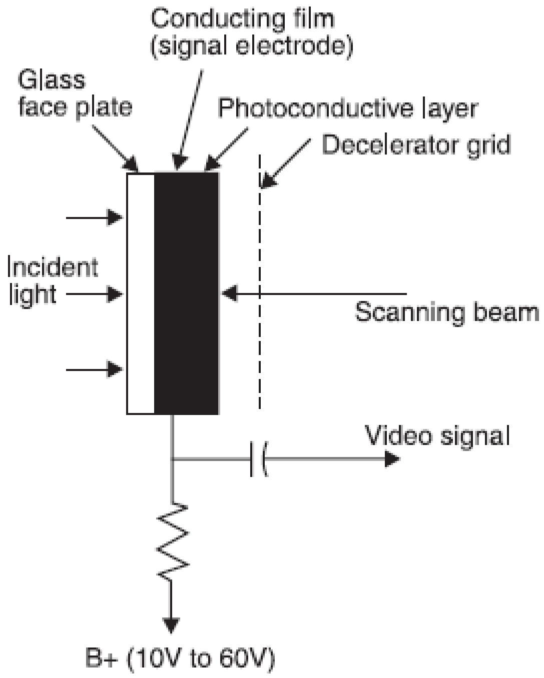


Fig. 30.30 Simplified diagram of the standard vidicon target plate. Output current is taken from the signal electrode.

Fig. 30.31 shows the *base diagram* for a typical vidicon. All connections except that made to the target ring are made at the base. In this case, there are places for eight pins equally spaced around the sealed exhaust tip. A short or missing pin serves as the *indexing system* for socket connections. *In most cases, the tube must be oriented in a particular way for optimum performance.*

Some vidicons made especially for portable cameras achieve beam focus by electrostatic means alone, Fig. 30.32. These employ several focus elements and use *higher accelerating voltages.*

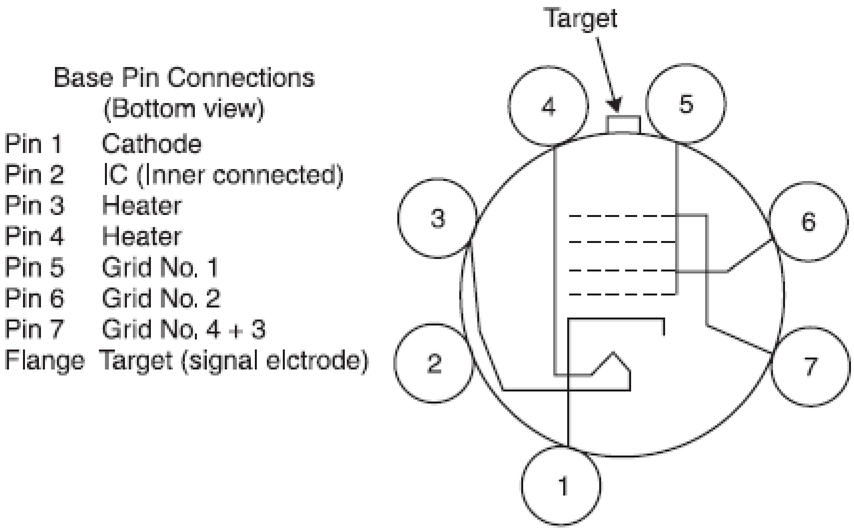
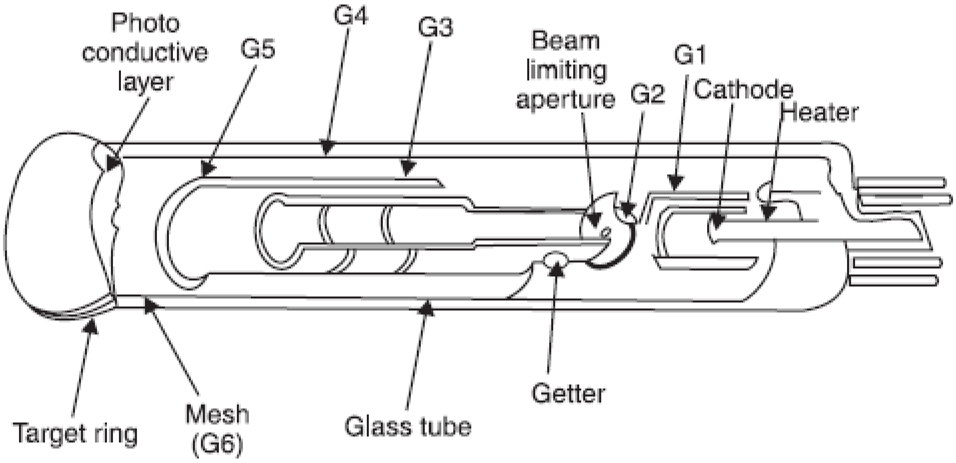
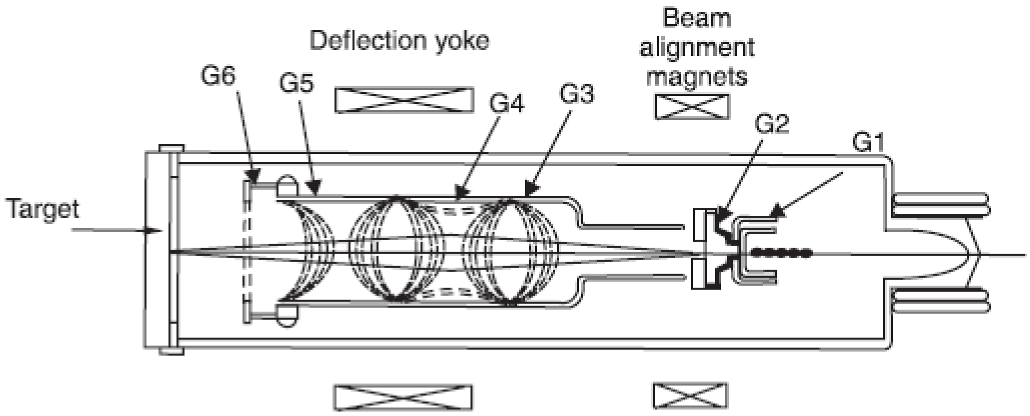


Fig. 30.31 Vidicon base diagram



(a) Construction details of H-V tube



(b) Electrostic focus lens and electrode details

Fig. 30.32 High voltage electrostatic gun uses a unipotential type lens with G3 and G5 at different potentials

Electrostatic lenses achieve a savings in electrical power in several ways. First, the current needed for magnetic focus coils is eliminated, as well as the power consumed by the series regulators that control focus current. Next, because the focus coil is eliminated, the space it occupies in the yoke assembly is eliminated. Magnetic reluctance in the gap between deflection coil poles is thereby reduced and less power is needed for deflection.

MONOCHROME PICTURE TUBE

The cathode ray tube or CRT—as the *picture tube* is commonly known as (shown in Fig. 30.33) is a large glass structure from which as much air has been evacuated as possible.

At the narrow end (the *neck)* of the tube is an assembly known as the *electron gun.* This comprises the elements which *emit and control* the electron beam just as a machine gun shoots and controls the direction of a stream of bullets.

The wide end (or the *face)* of the tube may be round or rectangular in shape. A thin coating or fluorescent material called a *phosphor* is deposited on the inside surface of the face. When a high speed beam, fired from the electron gun, strikes the phosphor, the face of the tube shows a tiny spot of light at the point where it strikes. Thus, *the screen converts the energy of moving electrons into light energy*.

All early television picture tubes had round screens. Since the TV picture is transmitted as a rectangle having an *aspect ratio* 4:3, there was much wasted space, or much wasted picture, if an attempt was made to use the entire tube screen-width. *Virtually all modern receivers utilise rectangular-faced tubes.*

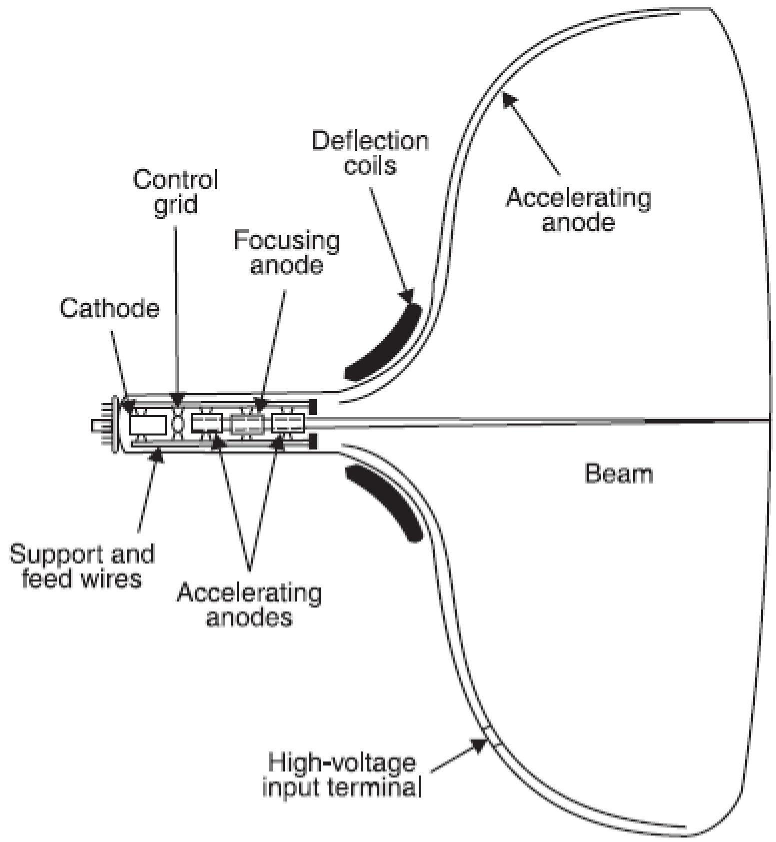


Fig. 30.33 The picture tube

Electrons are *liberated* from the filament or cathode of an ordinary vacuum tube. It is this stream of electrons that constitutes the plate current of the tube. Electrons released in exactly the same manner in the cathode ray tube form the television picture. The cathode of electron emitter is a small metal cylinder which is covered by an oxide coating. When the cathode is heated to a dull red by a heater wire located inside the cathode mounting, electrons are emitted by the cathode in large quantities. The electrons emitted by the cathode of an ordinary tube move in *all* directions while those coming from the cathode of the cathode ray tube are *emitted in a specific forward direction.*

The element that governs the *intensity* of the electron beam moving through the picture tube is the *control grid.* Unlike its counterpart in the common vacuum tube, it is cylindrical in shape and resembles a metal cap rather than a wire screen. This metal cap is called *Wehnelt cylinder,* Fig. 30.34. The front of the Wehnelt cylinder contains a small opening or *aperture* which acts as an outlet for the electron beam. This arrangement is necessary to achieve a *thin* electron beam.

Most picture tubes have a second cylindrical structure, placed adjacent to the control grid, as can be seen in Fig. 30.34. The principal function of the first anode is that of *accelerating* the electrons that emerge from the pinhole in the control grid structure. This accelerating action also causes some degree of beam forming, since the electrons are caused to increase speed as they pass through the apertures in the *first anode.* The potential applied to the first anode is in the vicinity of 250 to 450 volts. The position of this element is such that it causes the electrons coming from the grid to accelerate as a result of the electrostatic force of attraction between the negatively charged electrons and its positive potential.

Some picture tubes do not have a *first anode or accelerating grid*, as it is more appropriately called. In those tubes, acceleration is accomplished by high voltages applied to the *second anode*. The first anode initiates the *forming* of the electron beam because the positive charge on it draws electrons from the cathode. But, because we want the electron beam to *hit* the phosphor screen, we need another anode located closer to the screen. A second positive anode is added to the electron gun structure. This element usually carries anywhere from 8,000 volts to 20,000 volts (positive) depending on the *type* of picture tube.

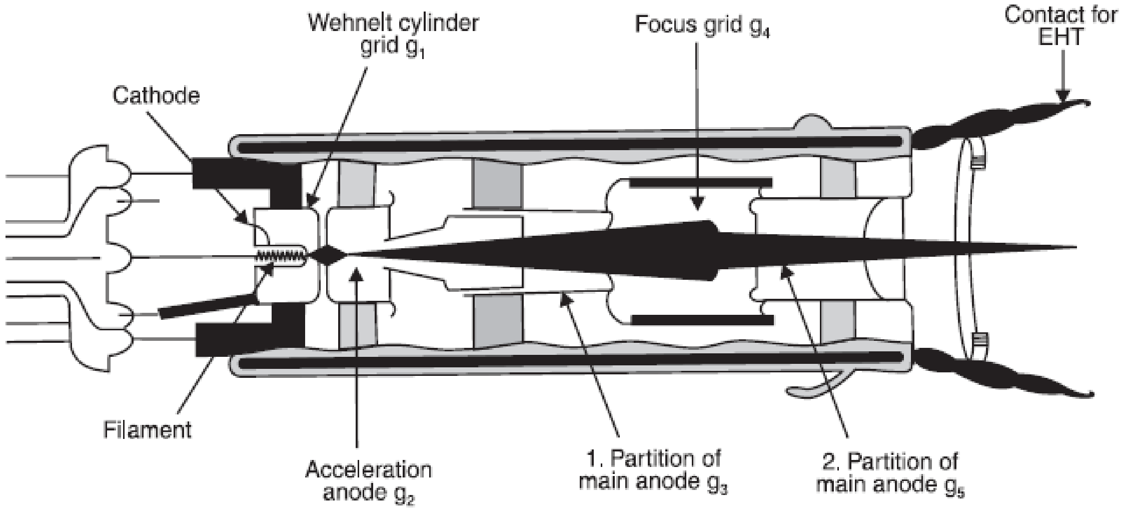


Fig. 30.34 Cathode ray tube construction

The *second anode* may consist of two distinct parts : (1) a cylindrical metallic structure adjacent to and following the first anode, and (2) a conductive coating inside the glass envelope which covers almost the entire area of the glass bell.

The conductive coating which forms a large part of the second anode is colloidal graphite deposit called *aquadag*. In addition to its accelerating action, the aquadag coating plays an important part in maintaining a thin electron beam, and assists in *filtering* the ripple from the high-voltage supply.

The glass picture tube sometimes has an *inner* conductive coating which forms part of the second anode and an *outer* coating of the same material that covers about the same area on the glass surface. *This electrical sandwich consisting of a nonconductor—the glass envelope—separating two conducting layers, forms a capacitor.* Although its capacitance is not very high (approximately 500 pf), it is large enough to serve as a filter-capacitor for the high voltage supply. The aquadag coating is connected to the chassis (ground) of the receiver as illustrated in Fig. 30.35. The inner coating is joined to the second anode and is connected to the high voltage terminal inside the receiver. *Thus, this capacitor is connected across the high-voltage supply which provides the potential needed for the second anode.*

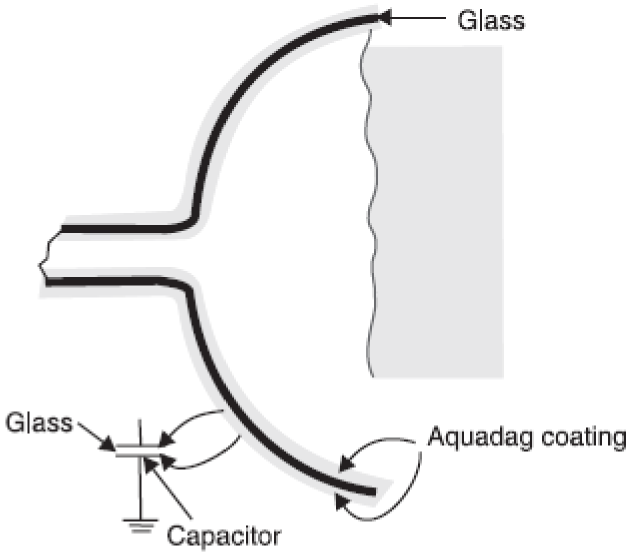


Fig. 30.35 High voltage filter capacitor

TELEVISION AS A SYSTEM

As shown in Fig. 30.36, television operates as a *system*, the station, and the numerous receivers within range of its signals. The station produces two kinds of signals, *video* originating from the camera or tape recording, and *sound* from a microphone or other source. Each of the two signals is generated, processed, transmitted separately, video as an *amplitude-modulated* signal, and sound as a *frequency-modulated* signal.

Considering such things as distance between stations, radiated power, geographic location, and topography, each station is assigned a particular *channel* to reduce the possibility of inter-channel interference. A station may operate in either of two *bands.* VHF (very high frequency) channels 2 through 13, extending from 54 to 216 MHz, or UHF (ultra high frequency) channels 14 through 83, extending from 470 to 890 MHz. Each channel occupies a *very wide bandpass* which can best be appreciated by realising that nearly seven radio broadcast bands would be required to accommodate even one TV station. The frequency allocations are in *numerical sequence* except for a couple of *gaps* where frequencies were previously assigned to other services.

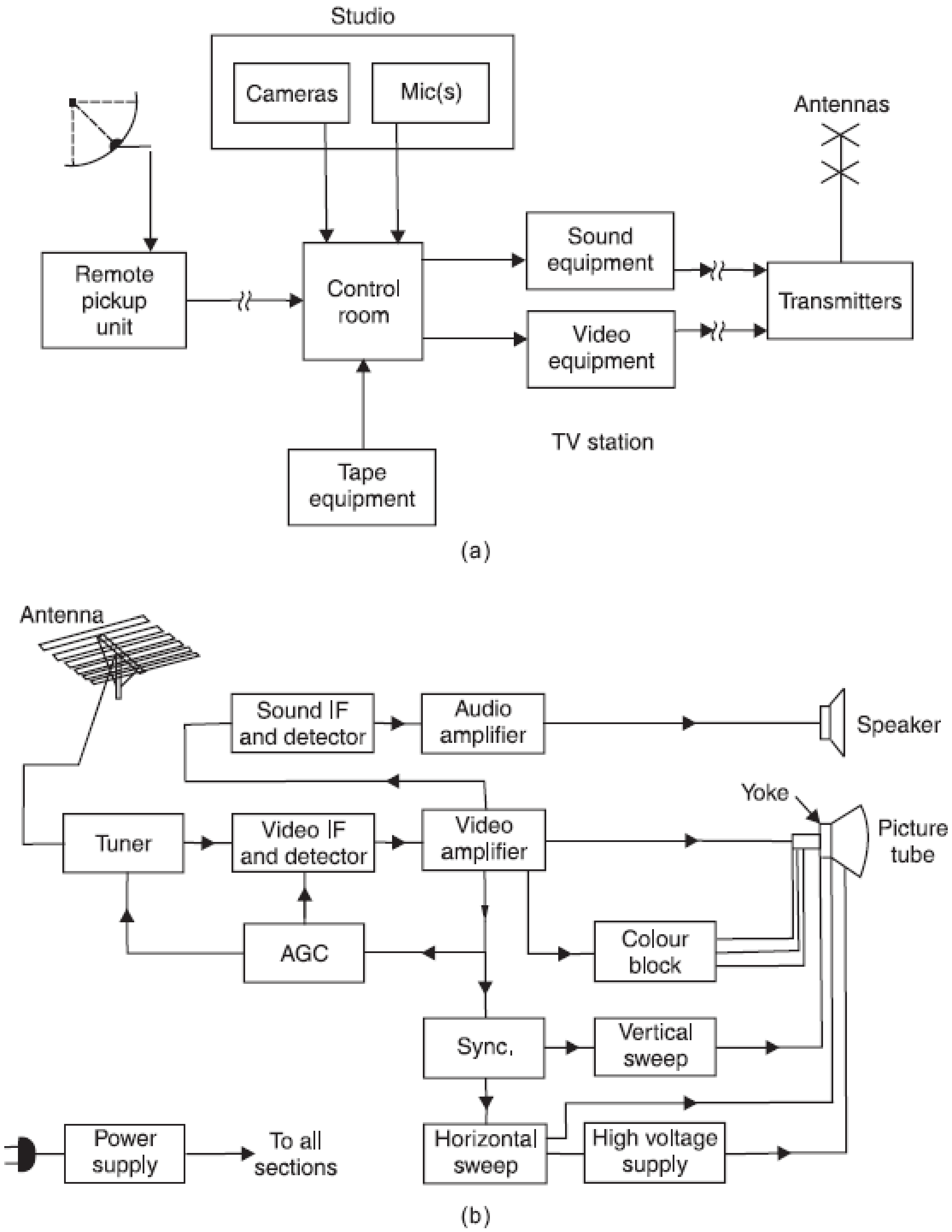


Fig. 30.36 Television as a system (a) transmitter and (b) receiver

AMERICAN 525-LINE TV SYSTEM

The transmitted television signal must comply with strict standards established by the Federal Communication Commission. The standards ensure that uniformly high quality monochrome and colour picture and sound will be transmitted. Standards are also required so that television receiver can be designed to receive these standard signals and thus provide a correct reproduction of the original picture and sound.

The details of the FCC standard are:

1. Each VHF and UHF television station is assigned a *channel* that is 6 MHz wide. The *composite video signal* must fit into the 6 MHz bandwidth. The composite signal includes the video carrier, one vestigial sideband for the video signal, one complete sideband for the video signal, the colour signals, the synchronising and blanking pulses, and the FM sound signal.
2. The *visual (picture) carrier frequency* is 1.25 MHz above the low end of each channel.
3. The *aural (sound) carrier frequency* is 4.5 MHz above the visual carrier frequency.
4. The *chrominance (colour) subcarrier frequency* is 3.579545 MHz above the picture carrier frequency. For practical purposes, the subcarrier frequency is expressed as 3.58 MHz.
5. One complete *picture frame* consists of 525 scanning lines (483 *active lines* actually produce the picture. The remaining lines are *blanked* out during the vertical retrace). There are 30 frames per second.
6. Each frame is divided into two *fields;* thus there are 60 fields per second. Each field contains 262½ lines, interlaced with the preceding field.
7. The scene is scanned from *left to right* horizontally at uniform velocity progressing *downward* with each additional scanning line. The scene is retraced rapidly (blanked out) from *the bottom to the top*, at the end of each field.
8. The *aspect ratio* of the picture is four units horizontally to three units vertically.
9. At the transmitter the equipment is so arranged that a decrease in picture light intensity during scanning causes an increase in radiated power. This is known as *negative picture transmission.*
10. The *video* part of the composite signal is *amplitude modulated.* The synchronising and blanking pulses are also transmitted by amplitude modulation and these pulses are *added* to the composite video signal.
11. The *colour signal* is transmitted as a pair of amplitude modulation sidebands. These sidebands effectively combine to produce a *chrominance signal* varying in hue or tint (phase angle of the signal) and saturation (colour vividness). Saturation corresponds to the amplitude of the colour signal. Colour information is transmitted by *interleaving* the colour signal frequencies between spaces in the monochrome video signals.
12. The *sound signal* is frequency modulated. As in FM broadcast systems the sound signal may also be produced by *indirect FM method.* The maximum deviation of TV is + 25 kHz.
13. Table 30.2 shows the *frequencies assigned* to the United States VHF and UHF channels.
14. According to FCC rules and regulations, *line 19* of the vertical blanking interval of each field is reserved for a special signal. This is the *vertical interval reference (VIR) signal*. The VIR signal is known as the chrominance reference. While originally intended for station use, it is also utilised by television receivers. At the station, the VIR signal is used as a reference to assure the correct transmitted hue and saturation of the colour signal. At the receiver, the signal and automatic circuits help to provide accurate colour reproduction. When this method is used, it eliminates the need for manual colour control adjustment.

Table 30.2 Television Channel Frequency Allocations

Channel Number	Frequency Limits (MHz) VHF Television	Picture Carrier (MHz) Station Frequencies	Sound Carrier (MHz)
2	54–60	55.25	59.75
3	60–66	61.25	65.75
4	66–72	67.25	71.75
5	76–82	77.25	81.75
6	82–88	83.25	87.75
7	174–180	175.25	179.75
8	180´186	181.25	185.75
9	186–192	187.25	191.75
10	192–198	193.25	197.75
11	198–204	199.25	203.75
12	204–210	205.25	209.75
13	210–216	211.25	215.75

UHF TELEVISION STATION FREQUENCIES

14	470–476	471.25	475.75
15	476–482	477.25	481.75
16	482–488	483.25	487.75
17	488–494	489.25	493.75
18	494–500	495.25	499.75
19	500–506	501.25	505.75
20	506–512	507.25	511.75
21	512–518	513.25	517.75
22	518–524	525.25	529.75
23	524–530	519.25	523.75
24	530–536	525.25	529.75
25	536–542	537.25	541.75
26	542–548	543.25	547.75
27	548–554	549.25	553.75
28	554–560	555.25	559.75
29	560–566	561.25	565.75
30	566–572	567.25	571.75
31	572–578	573.25	577.75
32	578–584	579.25	583.75
33	584–590	585.25	589.75

34	590–596	591.25	595.75
35	596–602	597.25	601.75
36	602–608	603.25	607.75
37	608–614	609.25	613.75
38	614–620	615.25	619.75
39	620–626	621.25	625.75
40	626–632	627.25	631.75
41	632–638	633.25	637.75
42	638–644	639.25	643.75
43	644–650	645.25	649.75
44	650–656	651.25	655.75
45	656–662	657.25	661.75
46	662–668	663.25	667.75
47	668–674	669.25	673.75
48	674–680	675.25	679.75
49	680–686	681.25	685.75
50	686–692	687.25	691.75
51	692–698	693.25	697.75
52	698–704	699.25	703.75
53	704–710	705.25	709.75
54	710–716	711.25	715.75
55	716–722	717.25	721.75
56	722–728	723.25	727.75
57	728–734	729.25	733.75
58	734–740	735.25	739.75
59	740–746	741.25	745.75
60	746–752	747.25	751.75
61	752–758	753.25	757.75
62	758–764	759.25	763.75
63	764–770	765.25	769.75
64	770–776	771.25	775.75
65	776–782	777.25	781.75
66	782–788	783.25	787.75
67	788–794	789.25	793.75
68	794–800	795.25	799.75
69	800–806	801.25	805.75
70	806–512	807.25	811.75
71	812–818	813.25	817.75
72	818–824	819.25	823.75
73	824–830	825.25	829.75

74	830–836	831.25	835.75
75	836–842	837.25	841.75
76	842–848	843.25	847.75
77	848–854	849.25	853.75
78	854–860	855.25	859.75
79	860–866	861.25	865.75
80	866–872	867.25	871.75
81	872–878	873.25	877.75
82	878–884	879.25	883.75
83	884–890	885.25	889.75

THE 625-LINE SYSTEM

The CCIR-B standard is adopted by our country. The details of the CCIR-B standard are:

1. Each VHF and UHF television station is assigned a *channel* that is 7 MHz wide. The *composite video signal* must fit into the 7 MHz bandwidth.
2. The *visual (picture) carrier frequency* is 1.25 MHz above the low end of each channel.
3. The *aural (sound) carrier frequency is* 5.5 MHz above the visual carrier frequency.
4. *The chrominance (colour) subcarrier frequency* is 4.43361875 MHz above the picture carrier frequency. For practical purposes, the subcarrier frequency is expressed as 4.43 MHz.
5. One complete *picture frame* consists of 625 scanning lines (585 *active lines* actually produce the picture. The remaining lines are *blanked* out during the vertical retrace). There are 25 frames per second.
6. Each frame is divided into two *fields;* thus, there are 50 fields per second. Each field contains 312½ lines, interlaced with the preceding field.
7. The scene is *scanned* from *left to right* horizontally at uniform velocity, progressing *downward* with each additional scanning line. The scene is retraced rapidly (blanked out) from *the bottom to the top*, at the end of each field.
8. The *aspect ratio* of the picture is four units horizontally to three units vertically.
9. At the transmitter, the equipment is so arranged that a decrease in picture light intensity during scanning causes an increase in radiated power. This is knwon as *negative picture transmission.*
10. The *video* part of the composite signal is *amplitude mlodulated.* The synchronising and blanking pulses are also transmitted by amplitude modulation and these pulses are *added* to the composite video signal.
11. The *colour signal* is transmitted as a pair of amplitude modulation sidebands. These sidebands effectively combine to produce a *chrominance signal* varying in hue and tint (phase angle of the signal) and saturation (colour vividness). Saturation corresponds to the amplitude of the colour signal. Colour information is transmitted by *interleaving* the colour signal frequencies between spaces in the monochrome video signal.

12. The *sound signal* is frequency modulated. As in FM broadcast systems, the sound signal may also be produced by indirect FM method. The maximum deviation of TV is + 25 kHz.
13. Table 30.3 shows the frequencies assigned to the VHF and UHF channels in India.

Table 30.3 Television Channel Frequency Allocations

Channel Number	Frequency Limits (MHz)	Picture Carrier (MHz)	Sound Carrier (MHz)
1	40–47	41.25	46.75
2	47–54	48.25	53.75
3	54–61	55.25	60.75
4	61–68	62.25	67.75
5	81–88	82.25	87.75
6	174–181	175.25	180.75
7	181–188	182.25	187.75
8	188–195	189.25	194.75
9	195–202	196.25	201.75
10	202–209	203.25	208.75
11	209–216	210.25	215.75
12	216–223	217.25	222.75
13	223–230	224.25	229.75
14	470–477	471.25	476.75
15	477–484	478.25	483.75
16	484–491	485.25	490.75
17	491–498	492.25	497.75
18	498–505	499.25	504.75
19	505–512	506.25	511.75
20	512–519	513.25	518.75
21	519–526	520.25	525.75
22	526–533	527.25	532.75
23	533–540	534.25	539.75
24	540–547	541.25	546.75
25	547–554	548.25	553.75
26	554–561	555.25	560.75
27	561–568	562.25	567.75
28	568–575	569.25	574.75
29	575–582	576.25	581.75
30	582–589	583.25	588.75
31	589–596	590.25	595.75
32	596–603	597.25	602.75
33	603–610	604.25	609.75
34	610–617	611.25	616.75
35	617–624	618.25	623.75
36	624–631	625.25	630.75
37	631–638	632.25	637.75

38	638–645	639.25	644.75
39	645–652	646.25	651.75
40	652–659	653.25	658.75
41	659–666	660.25	665.75
42	666–673	667.25	672.75
43	673–680	674.25	679.75
44	680–687	681.25	686.75
45	687–694	688.25	693.75
46	694–701	695.25	700.75
47	701–708	702.25	707.75
48	708–715	709.25	714.75
49	715–722	716.25	721.75
50	722–729	723.25	727.75
51	729–736	730.25	735.75
52	736–743	737.25	742.75
53	743–750	744.25	749.75
54	750–757	751.25	756.75
55	757–764	758.25	763.75
56	764–771	771.25	770.75
57	771–778	778.25	777.75
58	778–785	785.25	784.75
59	785–792	786.25	791.75
60	792–799	793.25	798.75
61	799–806	800.25	805.75
62	806–813	807.25	812.75
63	813–820	814.25	819.75
64	820–827	821.25	826.75
65	827–834	828.25	833.75
66	834–841	835.25	840.75
67	841–848	842.25	847.75
68	848–855	849.25	854.75
69	855–862	856.25	861.75
70	862–869	863.25	868.75
71	569–876	870.25	875.75
72	876–883	877.25	882.75
73	883–890	884.25	889.75
74	890–897	891.25	896.75
75	897–904	898.25	903.75
76	904–911	905.25	910.75
77	911–918	912.25	917.75
78	918–925	919.25	924.75

79	925–932	926.25	931.75
80	932–939	933.25	938.75
81	939–946	940.25	945.75
82	946–853	947.25	952.75
83	953–960	954.25	959.75

VESTIGIAL SIDEBAND TRANSMISSION

Because of the extensive bandwidth requirements of the video signal, it is desirable to make use of a *bandwidth saving technique.* The signal information is fully contained in each of the two sidebands of the modulated carrier; and *provided the carrier is present, one sideband may be suppressed altogether.* The single sideband transmission technique can reduce the bandwidth requirement to half, viz. 5 MHz. However, it is not possible to do this in the case of a television signal because a television signal also contains very low frequencies including even the dc information. It is impossible to design a filter which will cut off the unwanted band while passing the carrier frequencies and low-frequency components of the other sideband, without objectionable phase distortion.

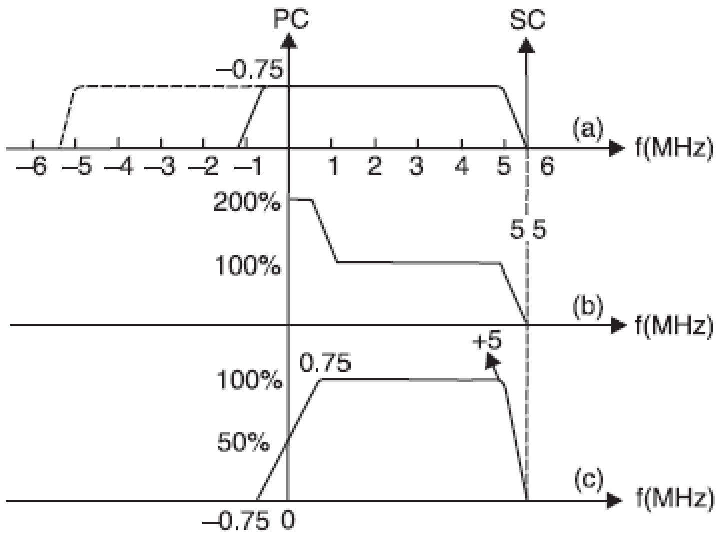


Fig. 30.37 Vestigial sideband channel characteristics (a) VSB channel (b) Videopower content and (c) Receiver response

As a compromise, a part of only one sideband is suppressed. The transmitted signal consists of one complete sideband together with the carrier, and a *vestige* of the partially suppressed sideband as shown in Fig. 30.37. In the vestigial sideband transmission system, *there is a saving in the bandwidth required*, and the filtering required is not so difficult to achieve.

EXERCISES

Descriptive Questions

1. What are the elements of a television system?
2. Explain the significance of scanning.
3. What are the different scanning methods?
4. Differentiate between linear scanning, sequential scanning and interlaced scanning.
5. Define the following terms: (a) Frame (b) Critical flicker frequency (c) Odd and even field (d) Trace and retrace (e) Active and inactive lines (f) Aspect ratio (g) Kell factor (h) Flicker (i) Vertical resolution (j) Horizontal resolution (k) Picture element (l) Video bandwidth (m) Test Card (n) Negative video polarity (o) Gamma.
6. What are the constituents of composite video signal?
7. Explain the working of a monochrome video camera with the help of a simplified diagram.
8. What are the characteristics of a camera tube?
9. Explain the working of a vidicon camera tube.
10. What is the difference between a camera tube and a picture tube?
11. With the help of a suitable sketch, explain the working of a picture tube.
12. Compare the 525 line American system with CCIR-B system. Which system is adopted in our country ?

Multiple Choice Questions

1. In the case of transmission of picture and sound (television) the fundamental problem is that of converting
   a. time variation into electrical information
   b. space variations into electrical information
   c. time and space variation into electrical information
2. Scanning may be defined as the process which permits the conversion of information expressed in
   a. space coordinates into time variations
   b. space and time coordinates into time variations only
   c. time coordinates into time variations
3. Aspect ratios of 4:3 are
   a. most pleasing to the eye
   b. least pleasing to the eye
   c. most fatiguing to the eye
   d. least fatiguing to the eye
4. Most practical television systems resort to
   a. linear scanning
   b. circular scanning
   c. nonlinear scanning
5. It is common practice to scan the image from
   a. left to right
   b. right to left
   c. top to bottom
   d. bottom to top
6. Most scenes have brightness gradations in a

a. vertical direction
b. horizontal direction
c. diagonal direction

7. The resolving capability of the human eye is about
    a. half-a-minute of visual angle
    b. one minute of visual angle
    c. one and a half minute of visual angle
    d. two minutes of visual angle
8. The television picture is traced by
    a. a spot of light
    b. a beam of light
9. Precision is essential in television but it is in the
    a. transmitting equipment and not receiving equipment
    b. receiving equipment and not transmitting equipment
10. The composite video signal locks in the frequency and the phase of
    a. frame oscillator
    b. line oscillator
    c. frame and line oscillators
11. The highest video frequency of a system depends on the
    a. number of lines
    b. rate of scanning
    c. both (a) and (b)
12. In the 625-line system, the sound carrier is
    a. amplitude modulated
    b. frequency modlulated
13. In the 625-line system the picture carrier is
    a. amplitude modulated
    b. frequency modulated
14. Transmitters in the centimetre wave range work almost exclusively on
    a. amplitude modulation
    b. frequency modulation

Fill in the Blanks

1. Scanning may be defined as the process which permits the conversion of information expressed in ____________________ and ____________________ coordinates into time variations only.
2. The detail which can be reproduced depends on the ____________________ of coverage by the scanning beam.
3. The process of scanning makes possible the use of a ____________________ transmission channel.
4. A varying scanning velocity for the pick-up and reproducing devices results in ____________________ output for ____________________ scene brightness.
5. It is common practice to scan the image from ____________________ to ____________________.

6. The repetition rate of flashes at and above which the flicker effect ____________________, ____________________, is called the critical flicker frequency.
7. The total number of pixels activated on the screen ____________________ the picture.
8. There is a degradation in vertical resolution due to ____________________ beam size.
9. The horizontal fly back interval occupies ____________________ of the line period.
10. The vertical flyback interval occupies ____________________ of the frame period
11. The vertical resolution of a system depends on the total number of ____________________ scanning lines.
12. The time during which the spot travels back to the top of the screen is called ____________________ period.
13. The method of scanning, where the odd lines and even lines are alternately scanned is called ____________________ scanning.
14. In television, the odd-line field and the even-line field make one ____________________.
15. The voltage representation of every line scanned is separated from the next by ____________________.
16. Each amplifying stage ____________________ the direction of video signal.
17. The video, sound, and control signals combine to make up ____________________ video signal.
18. The lowest frequency officially allocated to television by international agreement is ____________________.
19. The maximum deviation laid down for television sound is ____________________.
20. At least first ____________________ to ____________________ paired waves of each sideband must be transmitted.
21. The bandwidth of every link in the chain of FM transmission must be at least ____________________.
22. The ratio of the face plate illumination in foot candles to the output signal current in nanoamperes is called the ____________________ characteristic.
23. With camera tubes, the gamma value is generally ____________________.
24. With picture tubes, the gamma value is generally ____________________.
25. As nearly as possible, the camera tube should have the same spectral response as the ____________________ .
26. The formula for quantum efficiency is ____________________.
27. Quantum efficiency is a conveninet way to express the ____________________ of the image picture tube.
28. Lag on a television picture causes ____________________ to appear following rapidly moving objects.
29. Dark current refers to the current that flows through the device even when in total ____________________.
30. A short or missing pin in the base serves as the ____________________ system for socket connections.
31. All electrostatic lenses achieve a saving in ____________________ in several ways.

ANSWERS

Multiple Choice Questions

1.	(c)
2.	(b)
3.	(a) & (d)
4.	(a)
5.	(a) &(c)
6.	(a)
7.	(a)
8.	(b)
9.	(a)
10.	(c)
11.	(a)
12.	(b)
13.	(a)
14.	(a)

Fill in the Blanks

1.	space, time
2.	completeness
3.	single
4.	varying, constant
5.	left, right
6.	disappears
7.	reproduce
8.	finite
9.	17%
10.	6%
11.	active
12.	vertical retrace
13.	interlaced
14.	frame
15.	blanking interval
16.	inverts
17.	composite
18.	40 Mhz
19.	± 25 kHz
20.	five, seven
21.	200 kHz
22.	transfer
23.	1
24.	3
25.	human eye
26.	Q_{eff} = electrons/photons
27.	sensitivity
28.	smear
29.	darkness

30.	indexing
31.	electrical power

CHAPTER 31

COLOUR TV STANDARDS AND SYSTEMS

Colour television has considerably complicated the variables between different television systems. The systems used for television have been thoroughly standardised. There are three different incompatible systems for encoding signals. These three systems are used in different parts of the world. There are various steps in the processing of the video signal. These steps do not necessarily have to be the same in different countries, or not even in the same country, but the end result would always be the same.

DISPERSION AND RECOMBINATION OF LIGHT

If a stream of sunlight is made to pass through a small circular hole in one of the shutters, it will make a small circular patch on the opposite wall. On placing a triangular prism before the hole, an elongated coloured patch of light will be formed on the wall. This is called the *spectrum of light.* The colours formed are in the order shown in Fig. 31.1

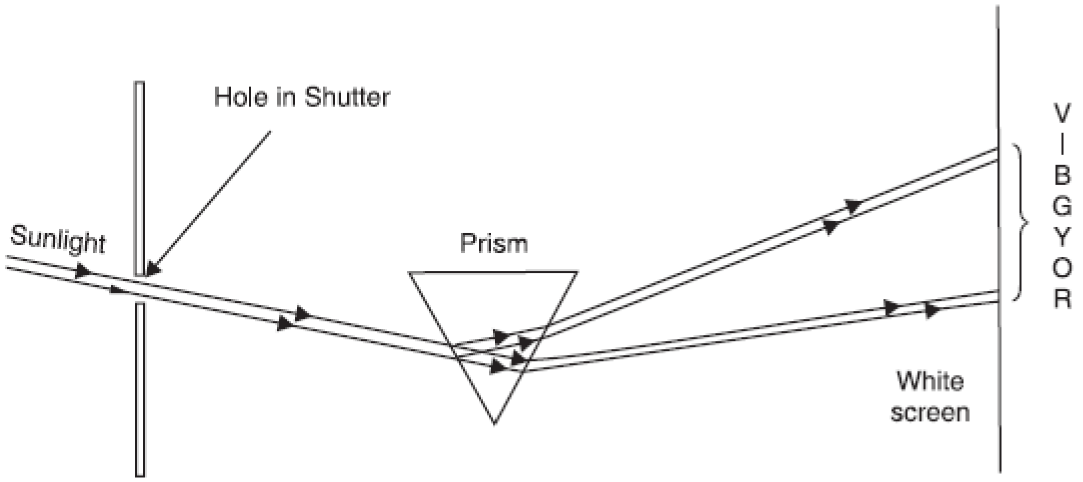


Fig. 31.1 Dispersion of light

White light consists of a mixture of seven different colours. The refractive index of glass is different for each colour, so that when white light falls on the prism, each colour in it is refracted at a different angle, with the result that *the colours are spread out to form a spectrum.*

When white light is incident on the prism it is refracted towards the base of the prism, *the violet being deviated the most and the red the least.* The separation of white light into its component colours by a prism is called **dispersion.** Strictly speaking there are many shades of each colour in the spectrum, each shade gradually merging into the next.

The colours of the spectrum may be recombined to form white light by allowing the spectrum to be formed on a row of small rectangular plane mirrors, Fig. 31.2. On adjusting the angle which the mirrors make with the incident light, so that all the mirrors reflect the light to the same place on the screen, the constituent colours of white light are recombined to form a white patch of light.

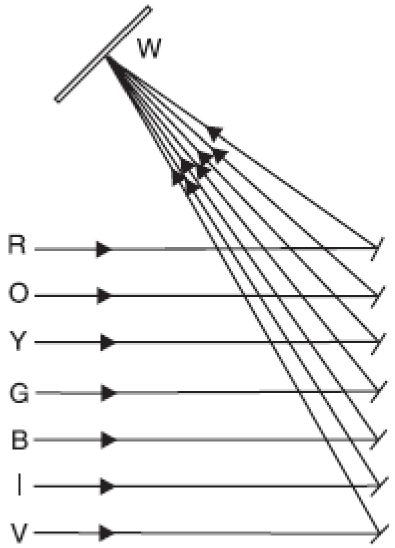


Fig. 31.2 Recombination of spectrum colours by mirrors

When white light falls on any particular body, then either all of the colours in the white light may be *reflected* from the body, when it appears white, or only some of them may be reflected while the others are *absorbed.* In the latter case, the body appears coloured. The energy of the light absorbed is generally covered into *internal energy* so that the body becomes slightly warmer. The *colour which the body presents to the eye is the colour of the light which it reflects.* Thus the leaves of plants appear green since they reflect green light and absorb other colours. *White* paper reflects all the colours of the spectrum, while black absorbs all of them, *Blackness is thus due to the absence of light of any colour.*

PRIMARY AND SECONDARY COLOURS

Colours which cannot be produced by mixing other colours are called primary colours. It is not found possible to produce either red, blue or green colours by mixing two other colours. For this reason red, green and blue are called primary colours.

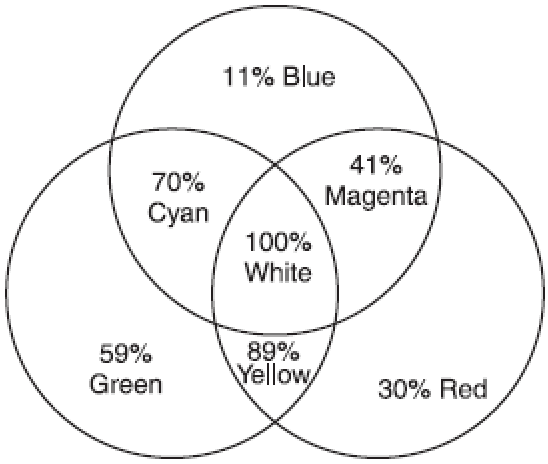


Fig. 31.3 Mixing of the three primary colours

A secondary colour can be produced by mixing other colours. Thus *yellow* colour can be produced by mixing red and green colours. *Magenta* colour can be produced by mixing red and blue colours. By the same token, *cyan* colour can be produced by mixing blue and green colours. This is shown in Fig. 31.3.

Two colours which give white light when added together are called *complementary colours*. The following facts can be verified.

$$\text{Primary colours} \quad \left\{ \begin{array}{ll} \text{Red + Green} & = \text{Yellow} \\ \text{Red + Blue} & = \text{Magenta} \\ \text{Blue + Green} & = \text{Cyan} \end{array} \right\} \quad \text{Secondary Colours} \qquad ...(31.1)$$

$$\begin{array}{l} \text{Primary and} \\ \text{complementary colours} \end{array} \quad \left\{ \begin{array}{ll} \text{Red + Cyan} & = \text{White} \\ \text{Green + Magenta} & = \text{White} \\ \text{Blue + Yellow} & = \text{White} \end{array} \right. \qquad ...(31.2)$$

In colour television, red, green and blue colours are chosen, and are called primary colours. When these colours are combined with each other in various proportions, a wide range of *hues and tints* (colour shades) are produced. These hues and tints are sufficient for presenting any colour picture. Also the *range of colours* produced by combining red, green and blue colours is *wider* than the range produced by combining any other colours.

The colours red, green, and blue are called *additive primaries* and are used when coloured light sources are blended to produce the required colour. Red, yellow and blue are called *subtractive primaries* and are used when a picture on print is viewed by reflected light from a white source.

ATTRIBUTES OF COLOUR

All primary colours as well as those produced by adding the primary colours *in suitable proportions* are characterised by three main features: hue, brightness and saturation.

Hue : The colour itself is called the hue and depends on the *dominant wavelength* of the light. By *adding* two or more of the primary colours many hues are produced.

Brightness : Each colour produces a certain amount of brightness. Brightness of a colour is determined by the *amount of light energy* contained in it. Light energy is measured in *lumens.*

Saturation : *The amount of white light* contained in a colour determines its saturation level. For instance, a highly saturated or deep red light will contain a much less amount of white light than a dull red light. *Saturation, thus denotes the degree of dilution of colour by white light.* More white light makes the colour dull or less saturated or vice versa.

Brightness of a colour is quite different from its saturation. *Brightness is an attribute of white light and colour light, whereas saturation is an attribute of colour light only.* It is possible to vary one while keeping the other constant. Colour TV receivers, therefore, have two different controls: brightness and saturation. When the *brightness control* is varied, the amount of light energy contained in the colour is varied. When the *saturation control* is varied, the amount of white light contained in the colour is varied.

LUMINANCE SIGNAL

The luminance signal can be obtained by adding together the signals representing the three primaries, R, *G* and *B*. However, the three voltages contributing the luminance signal must be taken in different amounts because *the human eye responds to each of the three primaries differently*. Calculations show that the luminance signal associated with whites of the picture should contain 30% red, 59% green and 11% blue.

$$E_Y = 0.30\,E_R + 0.59\,E_G + 0.11\,E_B \qquad \text{...(31.3)}$$

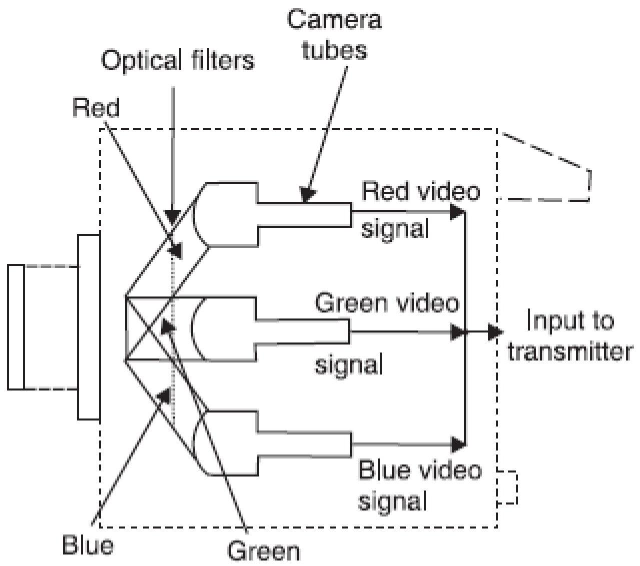


Fig. 31.4 The optics in a colour camera separate the scene into red, green and blue elements.

This equation, pivotal to colour television, needs some explanation, Suppose that a simple pattern made of a white stripe against a black background is projected onto photocathodes of the three pick-up tubes. The dichroic mirrors, Fig. 31.4, will split light from the white stripe into three colour components R, G and B. The gain of the video amplifiers can then be adjusted so that their output voltage is the same i.e.;

$$E_R = E_G = E_B$$

This is the *relative sensitivity* of a three tube camera.

The *luminance signal* essential to operation of a black-and-white TV, is produced by means of a *matrix*. The circuit of a simple matrix composed of four resistors (three *voltage dividers)* is shown in Fig. 31.5. If the values of R_1 R_2 and R_3 are chosen to be sufficiently high in comparison with R_{out}, the voltage dividers are *mutually isolated* so that the following voltages are developed across the resistor R_{out}:

$$\left.\begin{aligned} E_{Rout} &= E_R\,(R_{out}/R_1) \\ E_{Gout} &= E_G\,(R_{out}/R_2) \\ E_{Bout} &= E_B\,(R_{out}/R_3) \end{aligned}\right\} \qquad \ldots(31.4)$$

By setting the scale factors R_{out}/R_1 = 0.30, R_{out}/R_2 = 0.59 and R_{out}/R_3 = 0.11 the following luminance signal will be secured at the matrix output :

$$\begin{aligned} E_Y &= E_{Rout} + E_{Gout} + E_{Bout} \\ E_Y &= 0.30\,E_R + 0.59\,E_G + 0.11\,E_B \qquad \ldots(31.5) \end{aligned}$$

The following example will provide a better insight into the significance of equation 31.3 to colour television. Suppose two *blue and white* objects having the same intensity of radiation are to be televised. To an observer watching the two areas on a *colour reproducer*, both of them will appear different in luminance. In accordance with the above equation the luminance of the blue area will be equal to 11% of that of the white area.

On a black and white reproducer the areas will not appear coloured. However due to matrixing, the luminance of the blue area will still be equal to 11% of that of the white area because of the voltages E_R and E_G entering the equation for E_Y and coming from the red and green pick-up tubes, disappear and $E_Y = 0.11\ E_B$.

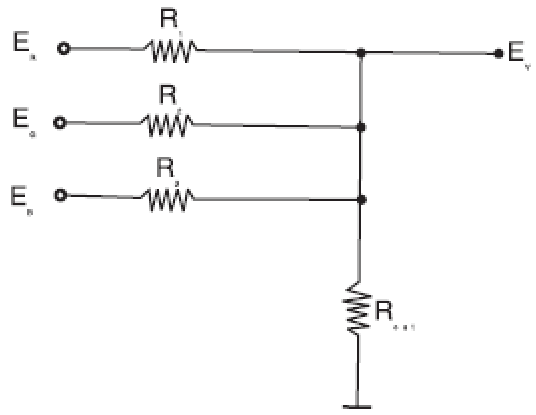


Fig. 31.5 Luminance signal

To keep the *bandwidth* needed for the television signal unchanged during the insertion of the luminance signal, one of the colouring signals, E_R, E_G or E_B should be dropped from the transmission. This may be done on the basis of equation 31.3. If the receiving terminal picks up two colour signals, say E_R and E_B in addition to the luminance signal, the third colouring signal E_G can be derived by matrixing. It follows from the equation for E_Y that

$$E_G = 1/0.59\ (E_Y - 0.11\ E_B - 0.30\ E_R) \qquad ...(31.6)$$

The gain of the succeeding stages should be adjusted so as to introduce a *correction factor* of 1/0.59. In this way, by the use of matrixing in the receiver, it is possible to derive four signals, E_Y, E_R, E_G and E_B from three input signals, E_Y, E_R and E_B.

CHROMINANCE SIGNAL

The generation of chrominance signal involves the *separation* of the luminance signal, E_Y from the natural colour signals, E_R, E_G and E_B. This is shown in Fig. 31.6. The natural colour signals E_R, E_G and E_B are developed across resistors R_1, R_2 and R_3 respectively. *Part of the voltage* across these resistors (0.3 V across R_1, 0.59 V across R_2, and 0.11 V across R_3) is tapped off. The full potential difference across R_1, R_2 and R_3 is simultaneously fed to the adders in stages 2, 3 and 4.

When a white scene is being televised, the three pick-up tubes will extract the three primary colours from the white light and develop *corresponding voltages* across R_1 R_2 and R_3. Since light falls on all the three pick-up tubes with the same intensity, these three voltages would be equal, *i.e*,

$$E_R = E_G = E_B = \text{(say) } 1 \text{ V}$$

Adder stage 1 will produce an output of:

$$0.30\ E_R + 0.59\ E_G + 0.11\ E_B$$

The sum of these voltages is the voltage E_Y corresponding to white light. As such, the output from adder stage 1, will be E_Y.

Thus $\qquad E_Y = 1 \text{ V}$

This output is routed through an L.P. filter (low pass filter) to an amplifier stage which is a conventional common-emitter amplifier, so that at its output we obtain the signal ($-E_Y$), whose value in this example will be –1 V.

The ($-E_Y$) signal is *applied simultaneously* to stage 2, 3 and 4. To these stages the full voltages E_R, E_G and E_B are also applied. In the case of white light each of these three voltages are each equal to 1 V. Thus each of the 2, 3 and 4 stages has two inputs :

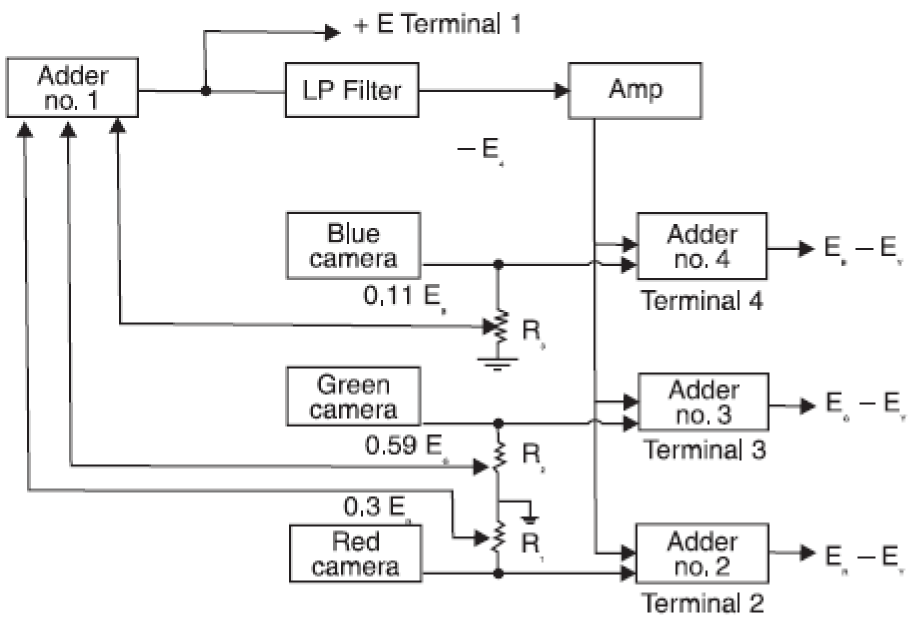


Fig. 31.6 Chrominance signal

$$-E_Y = -1 \text{ V}$$

and $$E_R = E_G = E_B = 1 \text{ V}$$

Obviously the output of each one of the adder stages 2, 3 and 4 will be zero. These *four outputs* are brought out from terminals 1 to 4 and routed to the TV transmitter after due processing.

This, then, is the process of televising a *white scene.* The luminance signal is available at terminal 1 for transmission and no signal is avialable at terminals 2, 3 and 4.

Let us next consider a *colour scene,* say a red scene. The red camera will generate the signal E_R. Let E_R, the potential drop across R_1 be 1 V. Part of the voltage (0.30 V) is tapped off and applied to terminal 1. Since only a red scene is being televised, there will obviously be no output from the blue and green cameras. Hence the input to terminal 1, will be:

$$E_R = 0.30 \text{ V}; E_G = 0 \text{ V and } E_B = 0 \text{ V}$$

Thus, the output designated as E_Y will be:

$$E_Y = 0.30 \text{ V}$$

This voltage is available at terminal 1 for being routed to the transmitter. Also, this voltage is *inverted* after passing through the LP filter and a stage of amplification. After inversion, it is applied to stages 2, 3 and 4. This inverted voltage will be.

$$-E_Y = -0.30 \text{ V}$$

Now the input to stages 2, 3 and 4 will be :

At terminal 2 $\quad E_R = 1$ V and $-E_Y = -0.30$ V

At terminal 3	$E_R = 0$ V and $-E_Y = -0.30$ V
At terminal 4	$E_R = 0$ V and $-E_Y = -0.30$ V

The output of the adder stages available at terminals 1 to 4 will be

At terminal 1	$E_Y = 0.30$ V
At terminal 2	$E_R - E_Y = 0.70$ V
At terminal 3	$E_G - E_Y = -0.30$ V
At terminal 4	$E_B - E_Y = -0.30$ V …(31.7)

The voltage available at terminal 1 is the *luminance signal* E_Y The voltages available at terminals 2, 3 and 4, designated as $E_R - E_Y$, $E_G - E_Y$ and $E_B - E_Y$ respectively are the *colour difference* or *chrominance signals.*

This, in brief, is the technique adopted in the TV transmitter for electrically separating the natural colour and luminance signals, resulting in the generating of colour difference or chrominance signals.

COLOUR PICTURE TUBE

The colour television camera *separates* the primary hues from the televised scene, Fig. 31.4, and the colour picture tube *recombines* them Fig. 31.7. This is comparable to the function of a mike (which picks up the sound from the program initiator and translates it into its electrical image) and louspeaker (which retranslates electrical image produced by the mike back into its original form for the listeners).

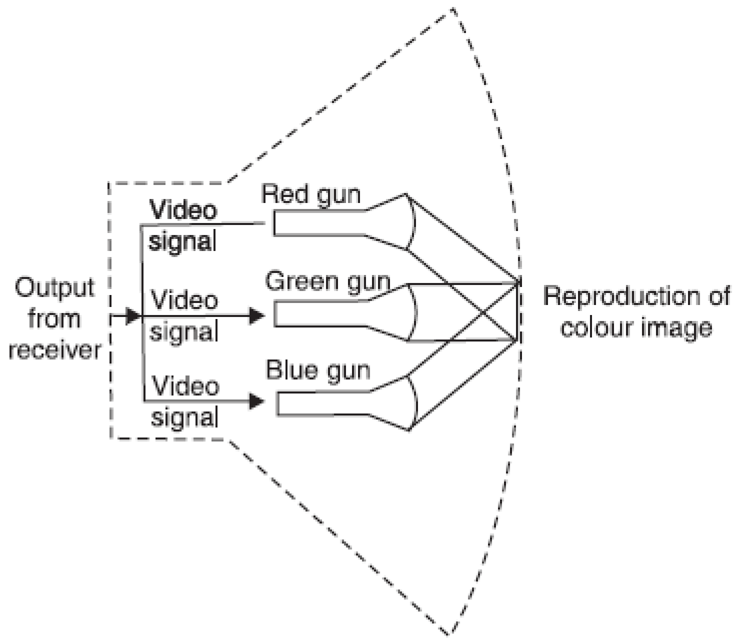


Fig. 31.7 The colour picture tube recombines the primary hues

The colour picture tube is basically a tube which has *three electron guns* in one picture tube envelope. The three guns are placed in the neck of the tube in a triangle *(delta).* One gun is called the *red gun,* the second the *blue gun* and the third the *green gun* (see Fig. 31.8). The combinations of these primary colours produce all the other colours, including white.

The *phosphor* on the screen or face plate is considerably different from the phosphor of monochorome picture tubes. Instead of a solid coating of one phosphor on the screen, each of the three phosphors is placed in dots along a horizontal axis, that is in one horizontal line, the three basic colour dots of phosphors will be placed in a triangle again and again, until the other side of the screen is reached. There are rows after rows of these triangles on the screen. They make up what is called a *phosphor dot pattern.* The colour tube, called the *trigun tricolour tube* also has a *shadow mask.* It ensures that the beams from the three electron guns hit their respective phosphor dots. This is shown in Fig. 31.8.

The colour tube is *scannded* in the same way as a black-and-white picture tube. If the picture being received is a *black-and white* picture all the three guns will be operating. If however, the picture being televised is *in colour*, the red gun will operate for red objects in the picture, green for green objects in the picture, and so on. If some other colour is required, then the *proper guns* will operate to mix the basic colours and produce the desired colour. The quantity of electrons hitting the phosphor dots is controlled by their respective control grids so as to produce any desired colour.

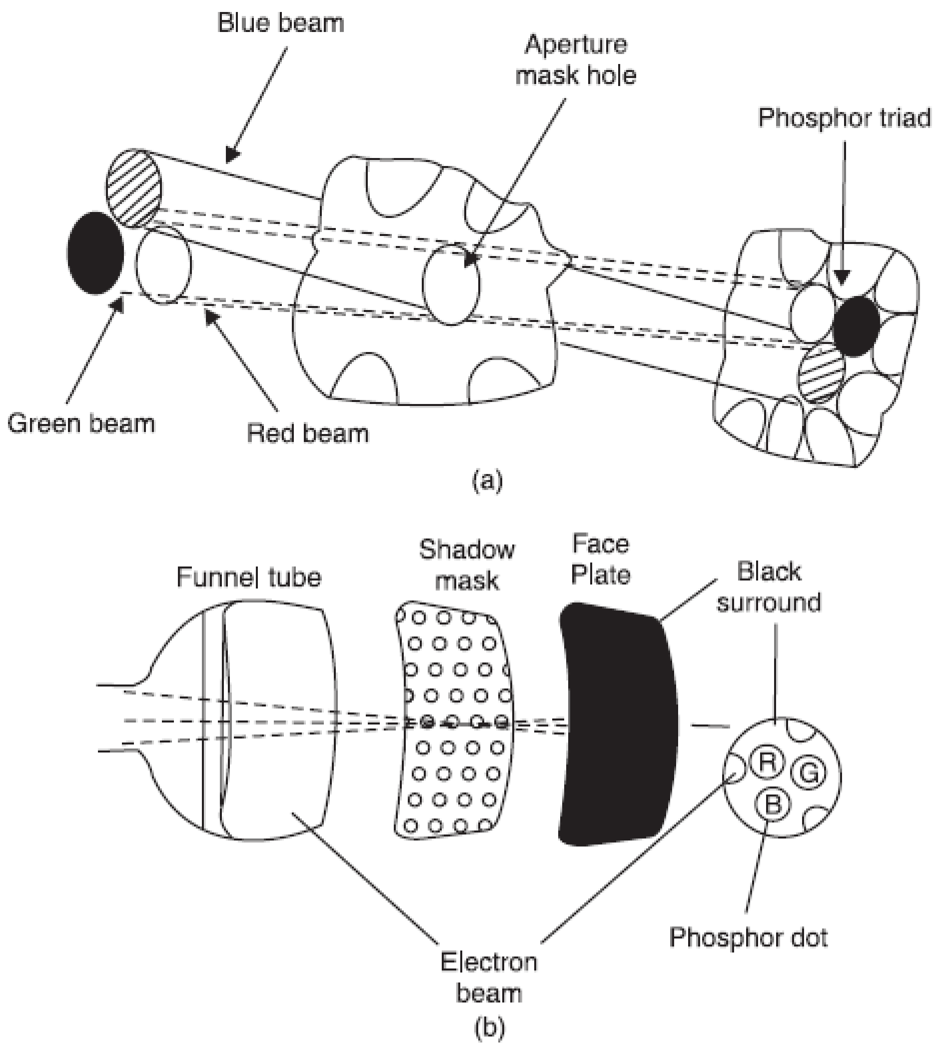


Fig. 31.8 Delta gun colour picture tube (a) Delta gun and (b) the shadow mask

DIFFERENCE BETWEEN A MONOCHROME AND A COLOUR PICTURE TUBE

The colour picture tube differs from the monochrome picture tube in the following four aspects:

1. It has *three guns* which provide *three electron beams*, one each for the three primary colours *(trigun, tricolour)*.
2. The screen of the colour picture tube is coated with *three different types of phosphor droplets* which separately emit red, green and blue light when the phosphors are bombarded by the high velocity electrons from the three guns.
3. The phosphors are embedded on the screen in a triangular dot pattern *(triad)*. Each triad has a group of three phosphor dots.
4. The *shadow mask* which is accurately mounted on the face about ½ inch behind the screen is the fourth distinguishing feature between a colour picture tube and a black-and-white picture tube.

DELTA GUN AND IN-LINE GUN

Shadow mask colour picture tubes are trigun, tricolour picture tubes with delta guns, Fig. 31.9. In picture tubes incorporating *delta guns*, each gun is positioned at an angle of 120° along the vertices of an equilateral triangle inscribed in a circle as shown in Fig. 31. 10(a). *Each gun can have the maximum possible size in the tube because of equal spacing between them. However, convergence adjustmens must correct for the guns in different planes.*

In *in-line guns* all the three guns are in one horizontal plane angled in towards the centre. Fig. 31.10(b).*For the same neck diameter, each in-line gun is smaller than each delta gun. With smaller guns, it is more difficult to obtain a high-intensity spot with sharp focus, but convergence is easier with in-line guns.*

There are three types of in-line gun tubes and they have completely replaced the delta gun tubes in colour TV receivers. Some picture tubes employ *one gun with three in-line cathodes*, as shown in Fig. 31.10(c). These cathodes produce three electron beams that are *focused and accelerated by a common gun.*

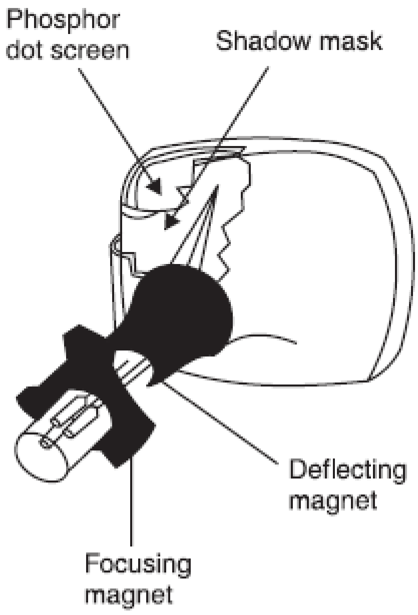


Fig. 31.9 Shadow mask colour picture tube

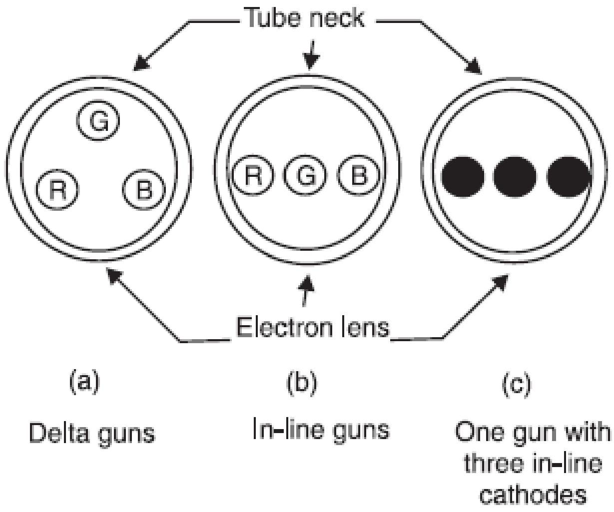
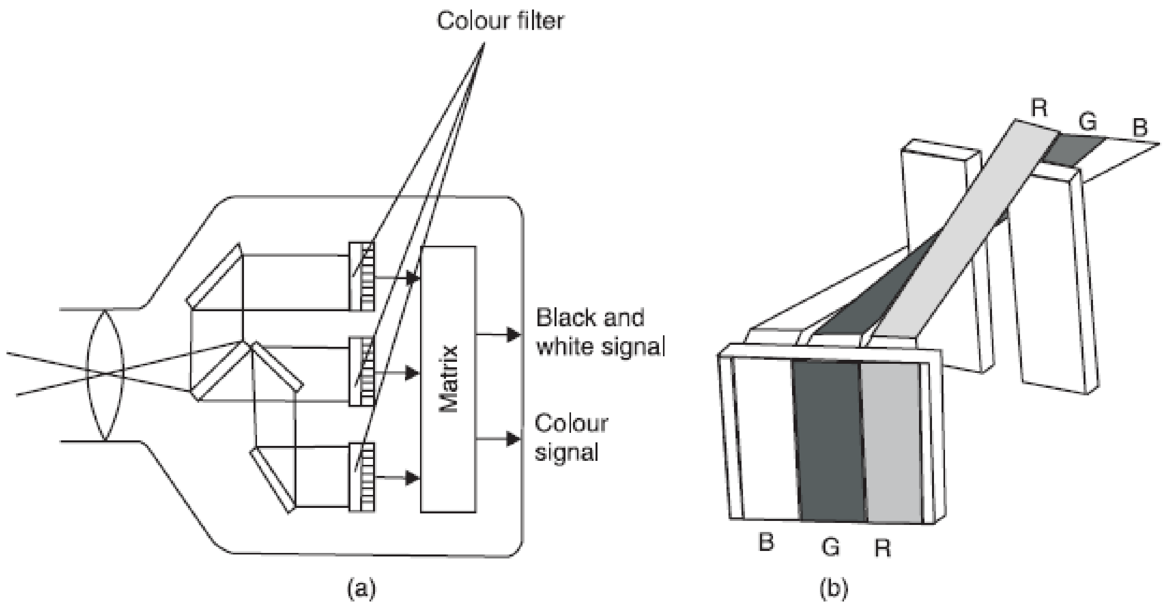


Fig. 31.10 Delta gun and in-line gun (a) Delta gun (b) In-line gun (c) One gun with three in-line cathodes

COLOUR TV CAMERAS

There are two general categories of colour TV cameras; one group uses a single (special) camera tube and the other group uses *three* camera tubes.

In the single-tube colour camera, Fig. 31.11, the tube face plate has a vertical stripe filter on the front of the tube and a *frequency separation technique* to extract the luminance and chrominance signals.



PLUMBICON CAMERA TUBE

The *plumbicon*, developed by Philips of Holland, is a small lightweight television camera tube that has *fast response* and processes high quality pictures at *low light levels.* Its *small size* and *low power operating characteristics* make it an ideal tube for solid-state TV cameras designed to serve a particular purpose. Modern colour television cameras are making widespread use of the plumbicon because of its simplicity and spectral response.

Functionally, the plumbicon is very similar to the standard vidicon. Focus and deflection are both accomplished magnetically. *The main difference between the plumbicon and the standard vidicon is the target.*

As shown in part (a) of Fig. 31.13 the inner surface of the glass face plate is coated with a thin transparent conductive layer of tin oxide (SnO_2). This layer forms the *signal plate* of the target. A photoconductive layer of lead monoxide (PbO) is deposited on the scanning side of the signal plate. These layers are specially prepared to function as three sub-layers. *Each layer has a different conduction mode.*

The tin oxide layer on the inner side of the face plate is a strong N-type semiconductor, commonly found in transistors. Next to this N-type region is a layer consisting of almost pure lead monoxide. This is an intrinsic semiconductor. The scanning side of the lead monoxide is doped to form a P-type semiconductor. Together these three layer's form a P-I-N junction diode.

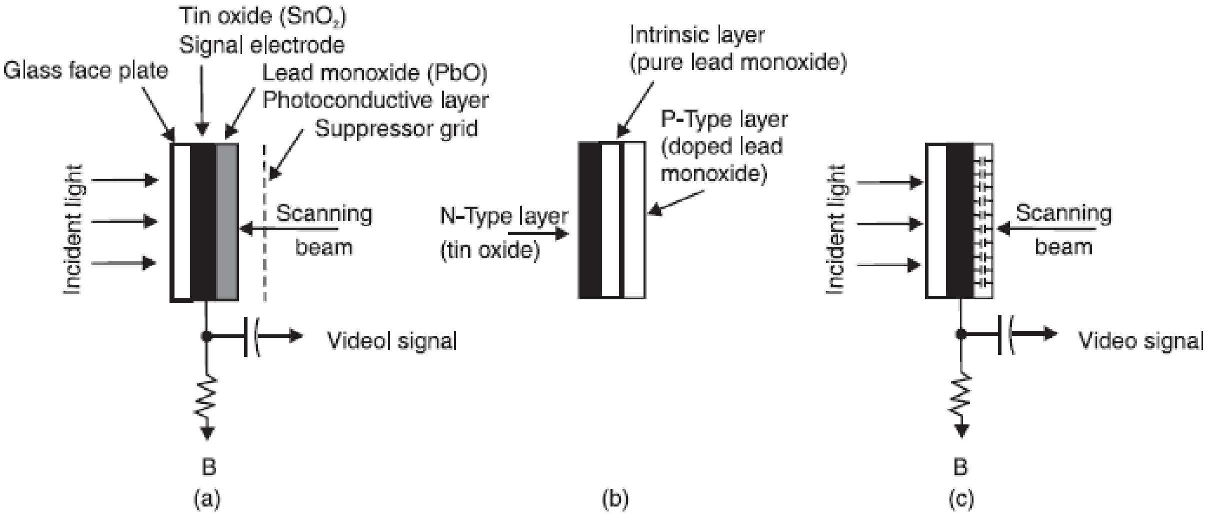


Fig. 31.13 A simplified diagram of the plumbicon target

Table 31.1 Typical Plumbicon Operating Conditions and Performance.

Operating Conditions (using coil unit AT 105)	
Cathode Voltage	0 V

Grid No. 2 Voltage	300 V
Signal Electrode Voltage	45 V
Grid No. 4 Voltage	500 V
Grid No. 3 voltage	300 V
Beam Current	150–300 nA
Monochrome Coil Assembly KV12	
Focus Current	120 mA
Line Current (P–P)	160 mA
Frame Current (P–P)	25 mA
Colour Coil Assembly At 1105	
Focus Current	40 mA
Line Current (P–P)	320 mA
Frame Current (P–P)	120 mA
Faceplate Temperature	20 to 45°C
	(68 to 113°F)
Blank voltage, peak-to-peak, grid No. 1	50 V
PERFORMANCE	
Dark Current	– 1.5 nA

Gamma of Transfer Characteristic (Gamma stretching circuitry is recommended)	0.95 ± 0.05
Spectral Response, max	500 nm
Cut-off XQ1427 G, B	650 to 850 nm
XQ1427 R	850 nm
Limiting Resolution	600TV Lines

The *photoconducitve target* of the plumbicon functions much like the photoconductive target in the standard vidicon. Light from the scene being televised is focused through the transparent layer of tin oxide onto the photoconductive lead monoxide. Each picture element charge takes the form of a *small capacitor* with its positive plate towards the scanning beam. The target signal plate becomes the negative side of the capacitor. When the low-velocity scanning beam lands on the charge element it releases enough electrons to *neutralise* the charge built up on the element capacitor. The scanning beam current through the external signal plate load resistor develops the video-signal output.

The spectral response of the plumbicon can be varied while it is being manufactured to suit almost any application. Since the tube gained wide popularity in colour television cameras, *it is available with a spectral response suitable for any of the primary colours.* The particular colour response of the tube is designated by the letter R (red), G (green) and B (blue) following its type number. When this type of tube is intended for *monochrome use*, no letter follows the type number.

COLOUR TV SYSTEMS

Colour television transmission and reception can be described as a system that reproduces colour images as depicted in Fig. 31.14, that is, a colour scene is *scanned* and *converted* into a corresponding video signal by a colour TV camera. In turn this signal is *processed* through the colour TV system Fig. 31.14(b) and is finally applied to a colour pictue tube. *A three-colour picture tube is used to reproduce the colour image.*

In its most general aspect a colour TV receiver can be regarded *as a conventional black-and-white receiver plus a chroma section and a colour picture tube*, as shown in Fig. 31.15. One of the prominent advantages of any of colour television system is its *compatibility.*

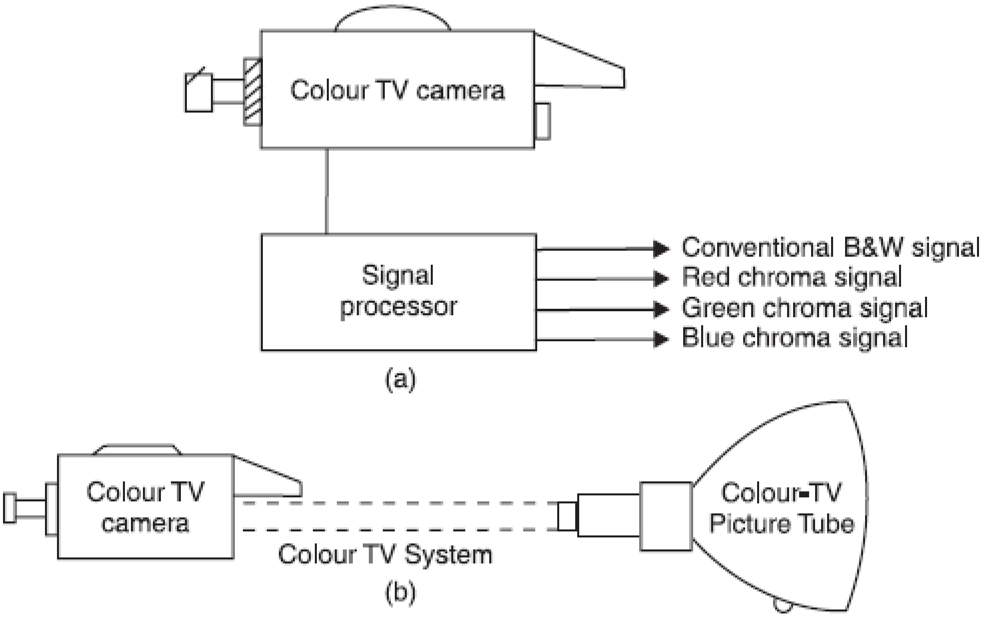


Fig. 31.14 (a) Complete colour signal comprises conventional black-and-white signal plus three chroma signals (b) Fundamental plan of the colour television system

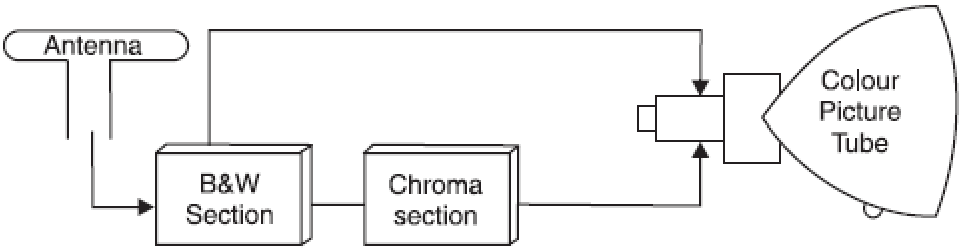


Fig. 31.15 Basic arrangement of a colour television receiver.

COMPATIBILITY CONSIDERATIONS

An *unmodified* black-and-white TV receiver cannot reproduce the image carried by a colour video signal in monochrome because the signal does not carry the component corresponding to a black-and-white picture. But it is possible by suitably adjusting a monochrome receiver, to reproduce one of the colour images (red, green or blue) in black-and-white. However, these cannot replace a black-and-white picture. For *compatibility* it is, therefore, necessary that in addition to information about colour, the transmitter of a colour television system should send a signal corresponding to black-and-white. This is usually called the *luminance signal* because the various areas of a black-and-white picture only differ in luminance.

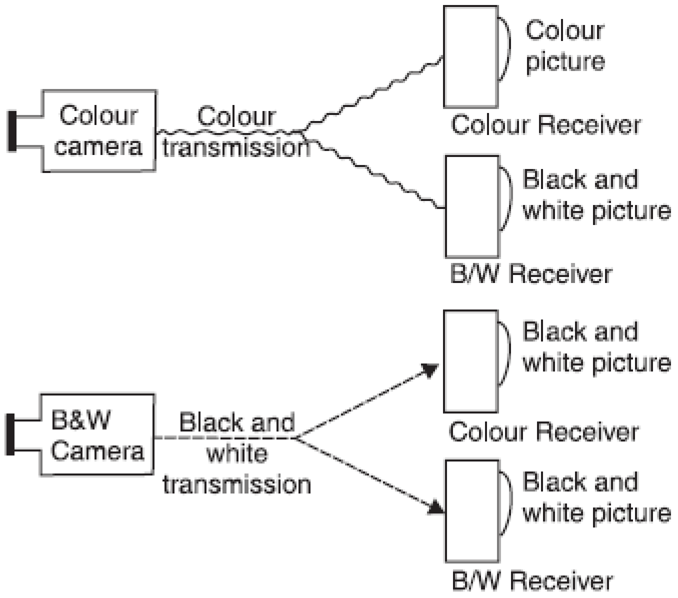


Fig. 31.16 Compatible system

The problem of making the colour and black-and-white programs compatible, now can be better understood. All that is required is to combine the two signals and transmit them together. One signal would contain the *luminance* or brightness information and the other the *chrominance* or colour information (hue and saturation). This is shown in Fig. 31.16.

For any two systems to be compatible, they must have the same scanning rate, colour subcarrier frequency and colour encoding technique. Presently, we have five different video systems: NTSC, PAL, PAL-M, PAL-N and SECAM. Some of these are *monochrome compatible* i.e. ; they can produce black-and-white pictures in the other compatible system. According to compatibility, the systems are divided into two groups with NTSC and PAL-M on one side and PAL, PAL-N and SECAM on the other.

Table 31.2 Monochrome Compatible Systems

525/60	625/50
NTSC	PAL
PAL-M	PAL-N
	SECAM

When both the brightness and colour information signals are transmitted a colour picture is produced on a colour receiver. But as a black-and-white receiver does not contain colour circuits, it can display only a *monochrome version* of the colour transmission. However, since all programs are not broadcast in colour, it is also essential that a colour receiver produce a satisfactory monochrome picture on a monochrome transmission. This is called, *reverse compatibility* (Fig. 31.17). The need for compatibility dictates the colour systems operate within the *same bandwidth* as the monochrome system and use the *same standards.*

If a compatible system is not adopted then the colour transmission will be available to only those viewers who are having colour receivers. The cost involved in the installation of a new set of transmitters *just* for the sake of colour programs would, no doubt, be prohibitive. These requirements are kept in view by all countries while switching over to colour transmission.

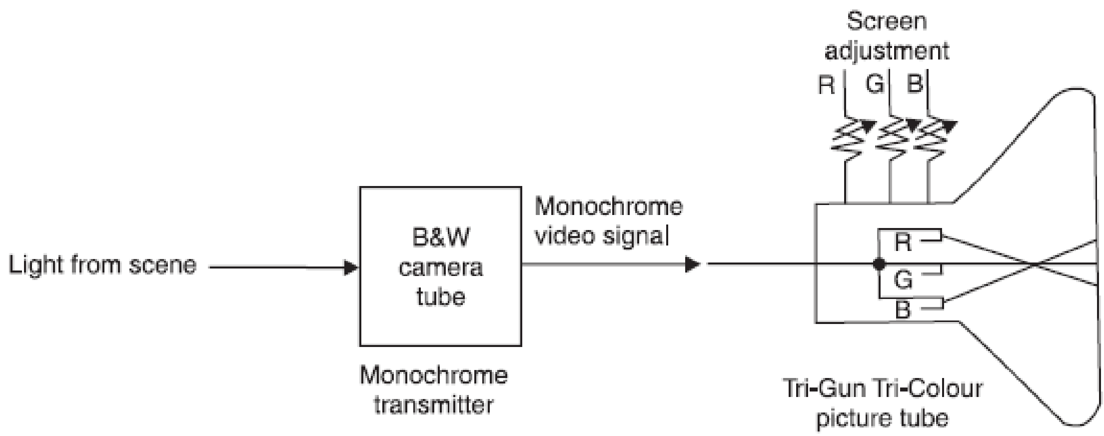


Fig. 31.17 A monochrome compatible system

THE NTSC SYSTEM

The world's first commercial colour television system was NTSC (National Television System Committee) introduced in the USA in 1953, and later adopted by Canada, Mexico and Japan. In the *NTSC system* the two colour difference signals, created by the subtraction of red and blue from the total signal, are not transmitted together, but with one a quarter of a cycle behind the other *(in quadrature).* The signals are then added together to form a single chrominance signal. When it reaches the receiver, the decoding circuits inside the TV break down the chrominance signal and *separate* it from its carrier wave. The two signals are then fed into a matrix which then *combines* them with the luminance signal to *recreate* the three original colour signals. These then create beams in three electron guns.

The main drawback in the NTSC system is that even *slight errors in the phase* between the colour difference signals produce errors at the decoding stage so that the set applied too much of one colour; hence the mnemonic for remembering the name *"Never Twice the Same Colour".* NTSC receivers have a hue control.

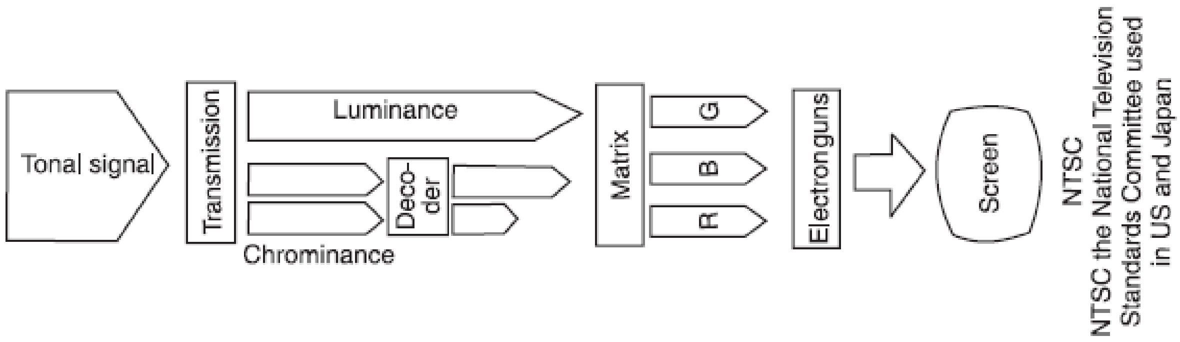


Fig. 31.18 The NTSC colour TV System

The NTSC system is a simultaneous system which uses quadrature modulation and is compatible with monochrome systems.

THE PAL SYSTEM

The PAL system, which is a *refinement* of the NTSC system, has been adopted in our country. *The PAL system aims to improve the colour picture.* The signals are transmitted in the same way, but the receiver *delays* the information on every line by means of an ultrasonic device for the exact time needed to compare it with the signal for the next line. If, for example, a certain line of the picture, as received, contains too strong a red signal, the system ensures the next line will be too low on red by reversing the polarity of alternate lines. *The final information passed to the picture tube is the average of the delayed first line and the corrected second line, so cancelling out the error. No hue control is necessary.*

The PAL (Phase Alternate Line) colour TV system may be said to stand midway between the NTSC and SECAM system. The PAL system is explained by the engineers in USA as "*Pay for Added Luxury*".

The PAL system is not a standard in its own right since the name refers only to the *technique* used to encode the colour information. *PAL standards* exist in both the 625/50 (erstwhile German Democratic Republic, UK, Australia and many other all using a 4.43 MHz colour subcarrier) and the 525/60 (the PALM system used exclusively in Brazil which has a 3.58 MHz colour subcarrier) scanning frequency standards. There is also a PAL-N version used exclusively in Argentina which is a *cross between the 625/50 scanning and 3.58 MHz* subcarrier.

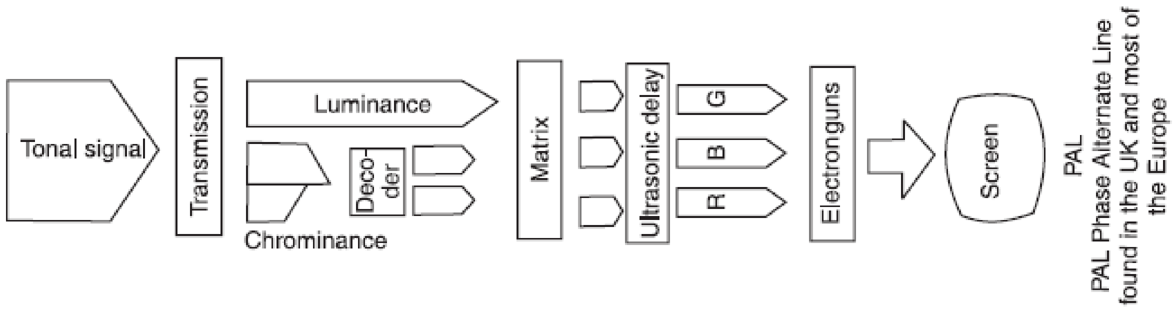


Fig. 31.19 The PAL colour TV system

PAL is a simultaneous compatible TV system using quadrature modulation. In contrast to SECAM, colour information in the PAL system is transmitted on every line.

Details of the different colour TV broadcasting systems are given in Table 31.3.

THE SECAM SYSTEM

In 1954, Henri De France, a French engineer, published the description of a new colour television system called the *"Henri De France System"*. However, due to a number of limitations this system did not find practical application. In 1956-1957, it gave way to a new design from which a whole family of systems has grown under the common name of SECAM (Sequential couleur a memoire). It is used in the erstwhile USSR and France.

The *SECAM system* has a different mode of transmission. The colour difference signals are not arranged a quarter of a cycle apart, but are kept separate by transmitting them on *alternate lines* of the picture. *Delay lines*, inside the receiver then *hold* up one set of signals so that they can be *recombined* to build a picture from alternate lines of signals from the original scan within the camera. The SECAM system does not, however, give good picture on a black-and-white receiver because it is difficult to separate the signals from their carrier.

In 1960-1961, the SECAM system was further modified to enable the colour difference signals to be frequency modulated onto a subcarrier, which considerably improved the performance of the system. *The use of frequency modulation, and the sequential transmission of chrominance signals are the major features of SECAM as distinct from the NTSC system.* In a SECAM receiver the chrominance signals are recovered on a time rather than a phase basis, thus rendering any synchronous detectors (as required in the NTSC system) unnecessary. *Frequency modulation has made the SECAM system insensitive to amplitude, frequency and phase distortions in the transmission circuit.*

SECAM is a compatible colour TV system. Its distinguishing feature is the fact that two colour-difference signals are transmitted sequentially on alternate lines, frequency modulated on a subcarrier while the luminance signal is transmitted on every line.

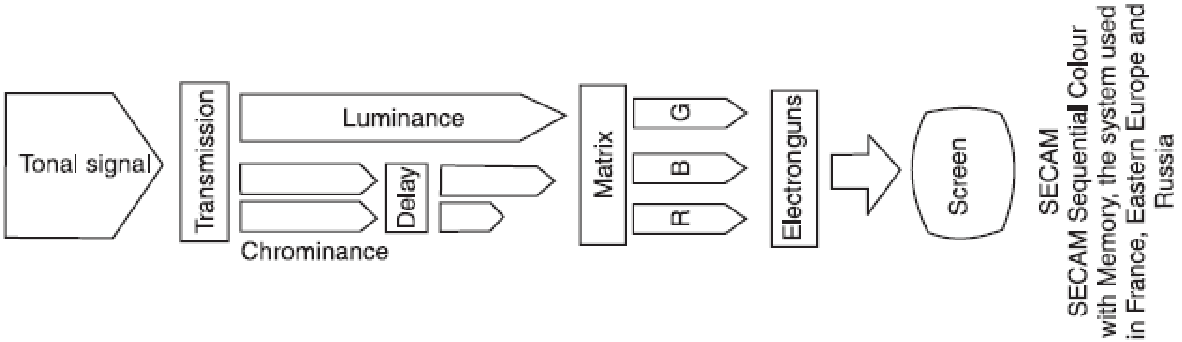


Fig. 31.20 The SECAM colour TV system

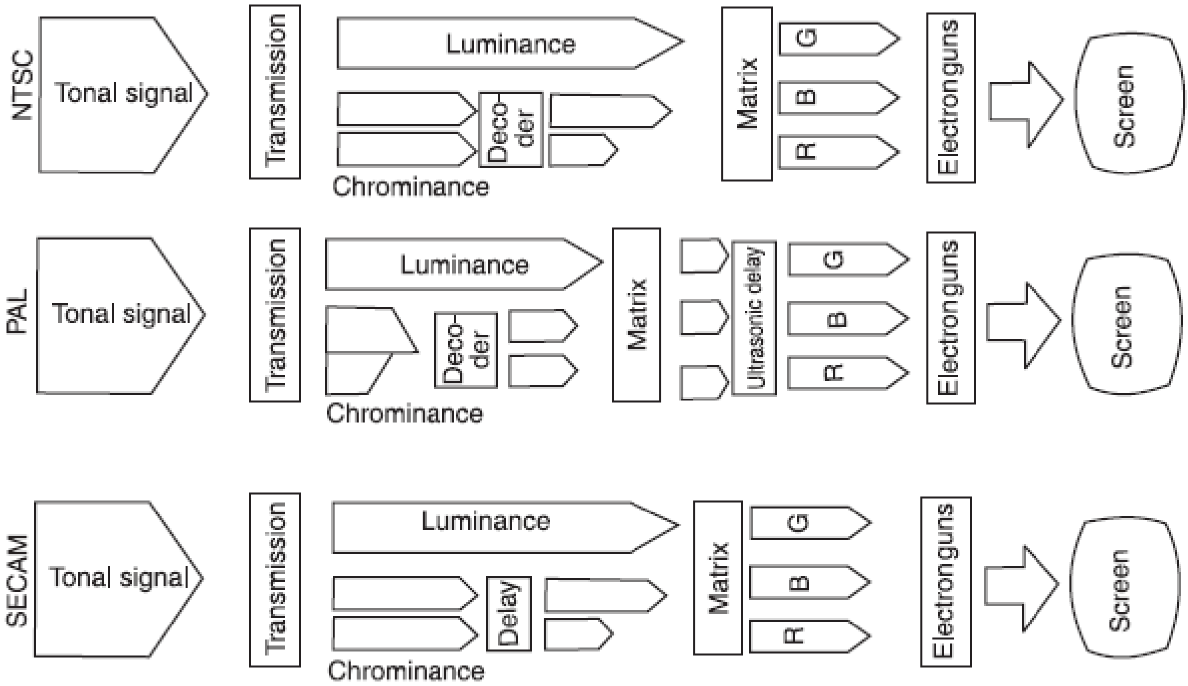


Fig. 31.21 Diagrammatic plans comparing and contrasting the three principal television broadcasting systems

THE SECAM SYSTEM

In 1954, Henri De France, a French engineer, published the description of a new colour television system called the *"Henri De France System"*. However, due to a number of limitations this system did not find practical application. In 1956-1957, it gave way to a new design from which a whole family of systems has grown under the common name of SECAM (Sequential couleur a memoire). It is used in the erstwhile USSR and France.

The *SECAM system* has a different mode of transmission. The colour difference signals are not arranged a quarter of a cycle apart, but are kept separate by transmitting them on *alternate lines* of the picture. *Delay lines*, inside the receiver then *hold* up one set of signals so that they can be *recombined* to build a picture from alternate lines of signals from the original scan within the camera. The SECAM system does not, however, give good picture on a black-and-white receiver because it is difficult to separate the signals from their carrier.

In 1960-1961, the SECAM system was further modified to enable the colour difference signals to be frequency modulated onto a subcarrier, which considerably improved the performance of the system. *The use of frequency modulation, and the sequential transmission of chrominance signals are the major features of SECAM as distinct from the NTSC system.* In a SECAM receiver the chrominance signals are recovered on a time rather than a phase basis, thus rendering any synchronous detectors (as required in the NTSC system) unnecessary. *Frequency modulation has made the SECAM system insensitive to amplitude, frequency and phase distortions in the transmission circuit.*

SECAM is a compatible colour TV system. Its distinguishing feature is the fact that two colour-difference signals are transmitted sequentially on alternate lines, frequency modulated on a subcarrier while the luminance signal is transmitted on every line.

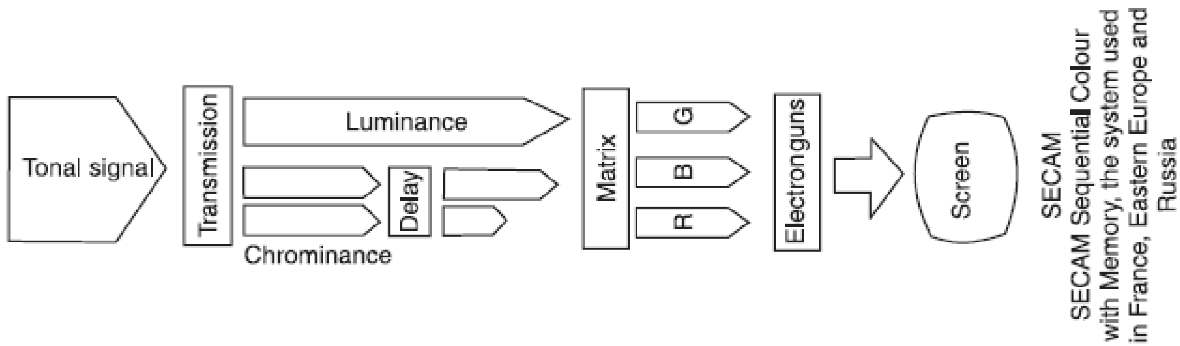


Fig. 31.20 The SECAM colour TV system

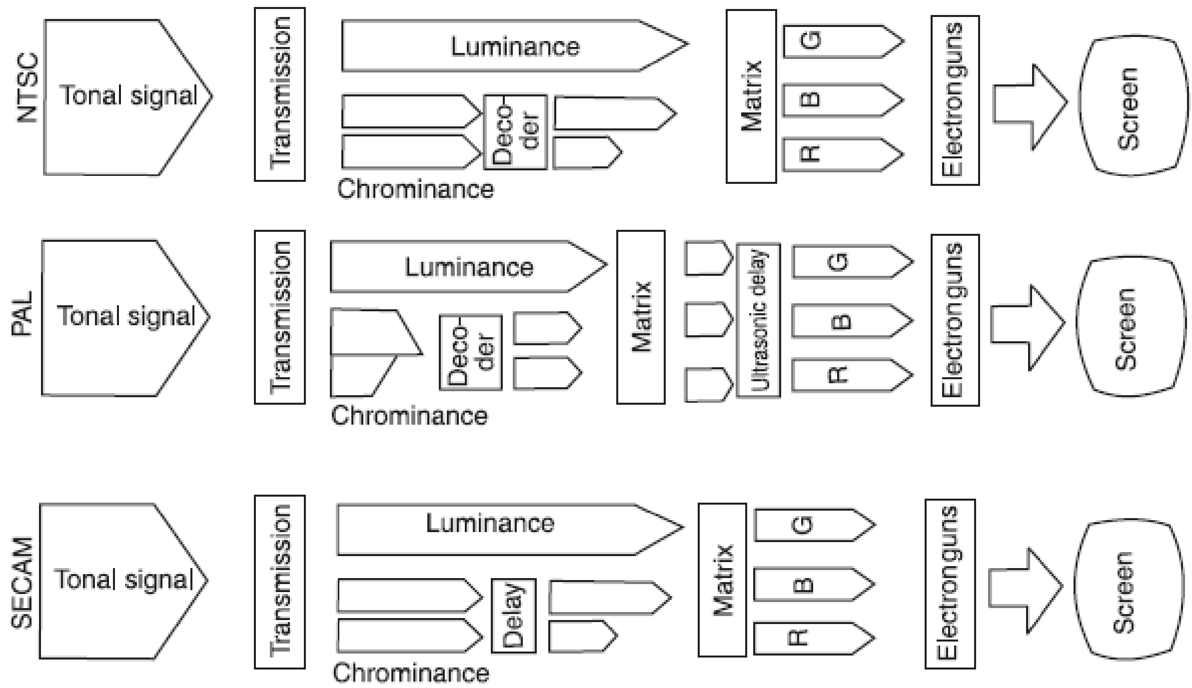


Fig. 31.21 Diagrammatic plans comparing and contrasting the three principal television broadcasting systems

BROAD CASTING OF TV PROGRAMS

The public television service is operated by broadcasting picture and sound from picture transmitters and associated sound transmitters in three main frequency ranges in the VHF and UHF bands. By international ruling of the ITU, these ranges are exclusively allocated to television broadcasting. Subdivision into operating channels and their assignment by location are also ruled by international regional agreement. The continental standards are valid as per the CCIR 1961 Stockholm plan. The details of the various system parameters are as follows.

Band	Frequency	Channel	Bandwidth
I	(41) 47 to 68 MHz	2 to 4	7 MHz
II	87.5 (88) to 108 MHz	VHF FM Sound	
III	174 to 223 (230) MHz	5 to 11 (12)	7 MHz
IV	470 to 582 MHz	21 to 27	8 MHz
V	582 to 790 (860) MHz	28 to 60 (69)	8 MHz
VI	11.7 to 12.5 GHz	superseded by satellite	
Special	68 to 82 (89) MHz`	2 (3) S	7 MHz
Channels	104 to 174 and	Channels	
Cable TV	230 to 300 MHz	S1 to S20	7 MHz

Types of Modulation

Vision : C3F (vestigial sideband AM)

Vestigial sideband ratios:

0.75 MHz/4.2 MHz = 1:5.6	for system M 525/60, 6 MHz
0.75 MHz/5.0 MHz = 1:6.7	for system B 625/50, 7 MHz
1.25 MHz/5.5 MHz = 1:4.4	for system I 625/50, 8 MHz

The saving of frequency band is about 40%; the ploarity is negative because of the susceptibility to interference of the synchronising circuits of early TV receivers (exception : positive modulation); residual carrier with negative modulation 10% (exception 20%).

Sound : F3E ; FM for better separation from vision signal in the receiver (exception : AM).

Sound carrier above vision carrier within RF channel, inversion at IF ; (exception : standards A, E and, in part, L).

Intermodulation distortion in TV transposers and TV transmitter with common vision-sound amplification and in Cable Television ; 20 : 1 : 0.2 for dual sound broadcast in the B/G standard *Channel bandwidth* are 5/6/7/8/14 MHz depending on standard, conventional values are 6/7/8 MHz and 14 MHz are still valid for a certain transition period.

SYSTEM PARAMETERS (FOR STANDARDS B/G)

Parameters	*Channel – 1*	*Channel – 2*
IRF sound carriers		
IF Frequencies	fvision + 5.5	fvision + 5.7421875
	MHz (+/–500 Hz), eqvt. to 353 fh	MHz (+/–500 Hz).
Vision/sound power ratio	13 dB	20 dB
Modulation	FM	FM
Frequency deviation max.	M +/–50 kHz	< + /–50 kHz
nominal value	+/–30 kHz	+/–30 kHz
Pre-emphasis	50 µs	50 µs
AF bandwith	40 to 15000 Hz	40 to 15000 Hz
Sound Modulation		
Mono	mono	mono
Stereo	(L + R)/2 = M	R
Dual sound mono	Channel A	Channel B
Identification		

Pilot carrier frequency	—	54.6875 kHz (+/–5Hz) Eqvt. to 3.5 fh)
Modulation	—	AM (with identification frequency)
Modulation depth	—	50%
Identification signal		
Mono	—	none
Stereo		(fh/133 =) 117.5 Hz
Dual sound		(fh/57 =) 274.1 Hz
Frequency deviation of transmitter carrier (due to pilot tone)		+/– (2.5 kHz +/–0.5 kHz)
Synchronisation		Pilot carrier and identification frequencies phase-locked with fh

The two *sound channels* arrive from the studio via radio link with 15 kHz bandwidth at the TV transmitter, where matrixing is performed for compatibility; (L + R)/2 for channel-1, R for channel-2. An additional sound modulator is used to modulate the second sound carrier with sound channel-2 and with the AM modulated pilot carrier.

The mode *identification* is transmitted in (data) line 16 (329) of a normal TV picture from the studio to the dual sound coder of the TV transmitter via the conventional TV lines *(i.e.* not the sound lines). From the 13 usable words of this data line the first two bits of word 5 are provided for mode identification in bi-phase code as follows :

Table 31.3 Colour TV Broadcasting Systems

Colour Systems	PAL				
Standard Broadcast Systems Items	B CCIR	D	G CCIR	I	N
Channel bandwidth (MHz)	7	8	8	8	6
Video bandwidth (MHz)	5	6	5	5.5	4.2
Frequency difference between audio and video	+ 5.5	+ 6.5	+ 5.5	+ 6	+ 4.5
Vestigial sideband (MHz)	0.75	0.75	0.75	1.25	0.75
Scanning lines	625	625	625	625	625
Field frequency (Hz)	50	50	50	50	50
Line frequency (Hz)	15625	15625	15625	15625	15625
Audio modulation method	FM	FM	FM	FM	FM
Video modulation method	AM (Negative)	AM (Negative)	AM (Negative)	AM (Negative)	AM (Negative)
Colour subcarrier frequency (MHz)	4.332618	4.332618	4.332618	4.332618	4.332618
Intermediate frequency (MHz)					
Audio	33.4	30.5	33.4	32.9	41.25
Video	38.9	37.0	38.9	38.9	45.75
	West Germany, Italy, Kuwait, Bahrain, Jordan, Algeria, Thailand, Indonesia, Australia, New Zealand, Holland, Norway, Switzerland, Sweden, Qatar, Malaysia, Pakistan Denmark, India	China	West Germany, Austria, Other European and African nations, * In Belgium and Yugoslavia vestigial sidebands are 1.25 MHz.	England (UHF) Hongkong (UHF) Ireland, South Africa	Argentina

Identification	Bit 1	Bit 2
Stereo	1	0
Mono	0	1
Dual sound	1	1
Fault	0	0

Vision/sound power ratio is 3 : 1/4 : 1/10 : 1/20 : 1, depending on the standard. Ratios of 5 : 1 and 10 : 1 are conventionally used ; 20 : 1 is used in Germany, *its advantage being energy saving.*

EXERCISES

Descriptive Questions

1. What are the constituents of white light?
2. Explain the difference in additive primaries and subtractive primaries.

3. What is the difference between a primary colour, secondary colour and complementary colour?
4. Differentiate between hue, brightness and saturation.
5. What is the need for a separate luminance signal?
6. Discuss the relative sensitivity of a three tube camera.
7. How are chrominance signals obtained?
8. What is the difference between a television camera tube and a picture tube?
9. How does the operation of a colour picture tube differ from that of a monochrome picture tube?
10. What are the differences between a delta-gun colour picture tube and an in-line colour picture tube?
11. Briefly explain the working of a colour camera tube.
12. How does a plumbicon differ from a vidicon? Discuss their relative merits and demerits.
13. What are the requirements of a compatible system?
14. How the different television systems have evolved?
15. Give the salient features of the television system used in our country

Multiple Choice Questions

1. The colour which the body presents to the eye is the colour of the light which it
    a. absorbs
    b. reflects
2. Primary colours
    a. can be produced by mixing other colours
    b. cannot be produced by mixing othe colours
3. The colours red, green and blue are called
    a. subtractive primaries
    b. additive primaries
4. The attributes of colour are
    a. phase
    b. hue
    c. amplitude
    d. saturation
    e. brightness
5. The signal essential to the operation of a black-and-white TV is
    a. chrominance signal
    b. luminance signal
    c. composite video signal
    d. control signal
6. In the NTSC system, the two colour difference signals are
    a. transmitted together
    b. not transmitted together
    c. transmitted in quadrature
    d. transmitted in anti-phase
7. Hue control is necesary in the
    a. PAL system

b. NTSC system
c. SECAM System
d. none of these

8. Colour information in which system is transmitted on every line
   a. PAL
   b. SECAM
9. Which television system is used in India

   a. PAL-N
   b. CCIR-B
   c. PAL-B
   d. PAL-G

Fill in the Blanks

1. White light consists of a mixture of _______________ different colours.
2. A _______________ colour can be produced by mixing other colours.
3. Hue depends on the _______________wave length of light.
4. The amount of white light contained in a colour determines its _______________ level.
5. Brightness of a colour is determined by the amount of _______________ contained in it.
6. Brightness is an attribute of _______________.
7. Saturation is an attribute of _______________ only.
8. The human eye responds to each of the three primaries_______________.
9. The receiving terminal picks up two colour signals, the third colouring signal _______________ is produced by _______________.
10. The generation of the chrominance signal involves the _______________ of the luminance signal from the _______________ signals.
11. The colour picture tube _______________ the primary hues from the televised scene.
12. The colour picture tube has _______________ electron guns in one picture tube envelope.
13. The delta gun picture tube has a _______________ mask.
14. The quantity of electrons hitting the _______________ dots is controlled by their _______________ control grids.
15. The screen of the colour picture tube is coated with _______________ different phosphor dots.
16. Some colour picture tubes employ one gun with three _______________ cathodes.
17. In the plumbicon camera tube, colours and deflection are both accomplished _______________.
18. The main difference between the plumbicon and the standard vidicon is the _______________.
19. Plumbicon camera tubes are available with a spectral response suitable for any of the _______________ colours.
20. For any two systems to be compatible, they must have the same, colour subcarrier frequency and colour _______________ techniques.

21. Monochrome compatible systems can produce _______________ pictures in the other compatible system.
22. The main drawback in the NTSC system is that even slight errors in the _______________ between the colour difference signals produce erors at the decoding stage.
23. Colour information in the PAL system is transmitted on_______________.
24. PAL standards exist in the _______________ system and _______________ system.

ANSWERS

Multiple Choice Questions

1.	(b)
2.	(b)
3.	(b)
4.	(b), (d) & (c)
5.	(b)
6.	(c)
7.	(b)
8.	(a)
9.	(b) & (d)

Fill in the Blanks

1.	seven
2.	secondary
3.	dominant
4.	saturation
5.	light energy
6.	white light and colour light
7.	colour light
8.	differently
9.	E_G, matrixing
10.	separation, natural colour
11.	recombines
12.	three
13.	shadow
14.	phosphor, respective
15.	three
16.	in-line
17.	magnetically
18.	target
19.	primary
20.	scanning rate, encoding
21.	black-and-white
22.	phase
23.	every line
24.	625/50, 525/60

UNIT – III

OPTICAL RECORDING AND REPRODUCTION

DISC RECORDING AND REPRODUCTION

When Edision first developed phonograph recording, he used vertical motions of the recording stylus to record sound vibrations. This came to be known as the 'hill and dale' method of recording. However, the 'hills' in the record groove had to lift the playback arm and cartridge. Since early arms and cartridges were relatively heavy, this motion caused excessive wear and tear. The record industry, therefore, adopted the lateral recording method, in which the needle moves from side to side in accordance with the sound vibrations. All commercial monophonic records for home use are laterally recorded, and monophonic pickups are designed accordingly. The year 1946 marks the start of the disc hi-fi. The binaural record and a special tonearm with which to play it appeared in 1952. In 1957, first stereo-disc was demonstrated and the moving-magnet stereo pickup was invented.

MAKING THE TAPE

The live performance is recorded on tape, Fig. 8.1. Often, each instrument and a singer is given a separate microphone and a separate channel. Professional tape recordings may have as many as 32 channels so that the sound of each instrument can be individually monitored - balanced, adjusted, and enhanced. Different channels can be recorded at different times.

The placement of microphone is very important and the optimum position are decided with experience.

The multiple channels recorded in the studio are mixed down to stereo before the record is cut. At the mixing stage, record producers make many decisions that determine the end sound quality of a performance.

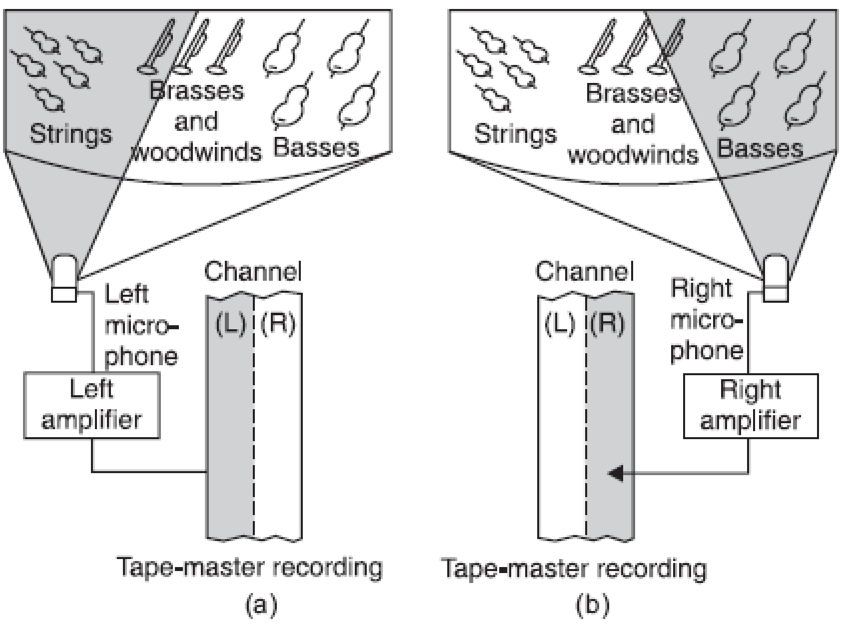


Fig. 8.1 Tape-master recording

In the *standard recording procedure* Fig. 8.2, the microphones feed the signals to preamplifiers. From there the signals go to amplifiers, then to a tape recorder. In recent innovations with *noise reduction systems* the process is much more complex but offers greater reduction of noise.

The tapes run at speeds of 15 or 30 inches per second (ips) to ensure high quality and accurate response.

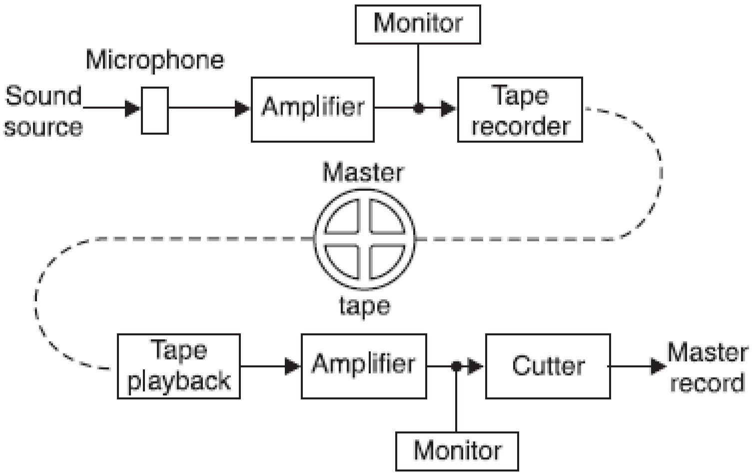


Fig. 8.2 Standard recording procedure

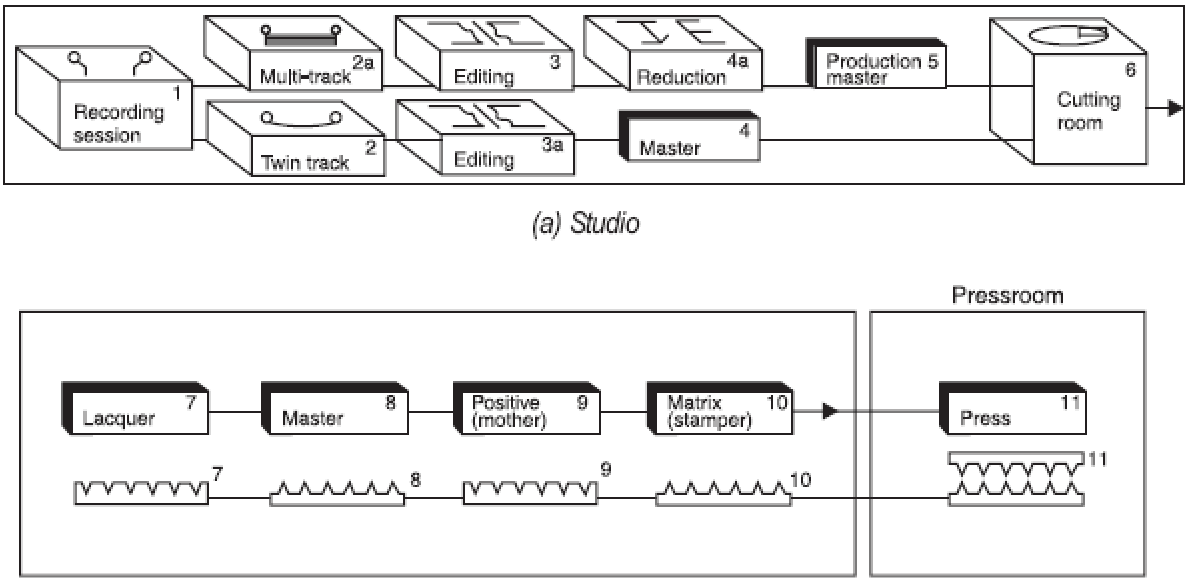
MAKING THE RECORD

When the master tape is played, its magnetic patterns are converted into electrical signals. These are fed through amplifiers to a *cutter head* on a lathe which converts these signals into mechanical motion and causes a cutting stylus to move as dictated by the signals on the tape. Thus, an incredibly complex groove is inscribed into a master disc. Moulds are then made from which the records are stamped.

PROFESSIONAL RECORD MANUFACTURE (FIG. 8.3)

1. The *recording session* may be in a studio or an external location. A simple two-microphone balance may be used, or a multi-microphone set-up.
2. Multi-mix as well as *two-mix sessions* can be recorded directly onto twin-track 1/4 in tapes.
2a. *Multi-mix sessions* are more commonly recorded on multi-track machines (1 or 2 in tape : 8, 16, 24 or 32 tracks).
3. and 3a. The best takes are edited together to produce a complete performance.
4. An edited twin-track tape is called a master and will usually be copied for safety.
4a. A multi-track tape is copied *via* a mixing desk by a process known as *reduction*, to make…

5. ….*twin-track production master*.
6. The twin-track tape goes to the cutting room for *transfer to lacquer*.
7. The lacquer is *sprayed with silver* to make it conductive.
8. It is put into an electroplating bath and a *nickel master* is grown onto the silver coating. The lacquer is *discarded* leaving the master which is now a *mirror image of the master*.
9. By a further plating operation a *positive* (or *mother)* is grown. The master can be re-used.
10. From the positive a *matrix* (or *stamper)* is grown and this has *protruding grooves* like the master.
11. A *pair of stampers* goes to the press to make a record.
12. The main *constituents* of the vinyl are a vinyl copolymer, a stabiliser, a lubricant and a pigment. The latter is usually carbon black, to produce the *traditional record colouring*.
13. The materials are *compounded* and fed to the press by extrusion or as cakes of small dice-like particles; from these the *final product* is pressed.
14. Manual semi-automatic and fully-automatic presses are all used in record manufacture today.



(a) Studio

(b) Electroforming (Plating)

(c) PVC manufacture

Fig. 8.3

STEREO PICKUP TECHNIQUES

The most general type of stereo pickup technique is called *time-intensity pickup.* This means that the signals in the two stereo channels will vary in both intensity and time according to the difference in direction and in distance of the sound source from the two microphones.

The stereo effect can be obtained by picking up the sound with two microphones spaced a certain distance apart and located in front of the sound source. If they are too close the stereo effect is lost. Experiments have led to the conclusion that *there is a definite optimum relation between the microphones separation and their distance from the source. A similar relation exists between the listeners' speakers and their distance from him.* This is illustrated in Fig. 8.4. In practice the *angle* at the listener should not be less than 30 degrees nor exceed 45 degrees. This relation keeps the distance of the listener *from* the speakers approximately the same as the distance between the speakers and ensures sufficient time intensity variation to provide stereo effect.

A good rule to follow is to spread the microphones as far apart as you can and still pickup an appreciable amount of sound from the middle portion of the source. Usually the pattern will be such that they can be placed along the sides of a triangle (see Fig. 8.4).

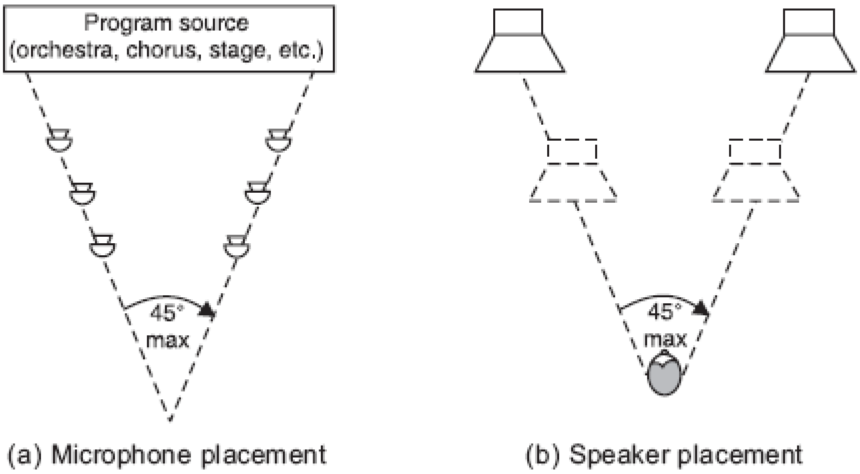


Fig. 8.4 Microphone and speaker placement

Yet another method, referred as *intensity-difference system*, is sometimes used in stereo recording. In this system, the microphones are not spaced; however, they have a directional characteristic as shown in Fig. 8.5. *The stereo effect is obtained by proper orientation of the microphones.* One method is to mount the microphones at 90 degrees to each other and 45 degrees to the centre of the source.

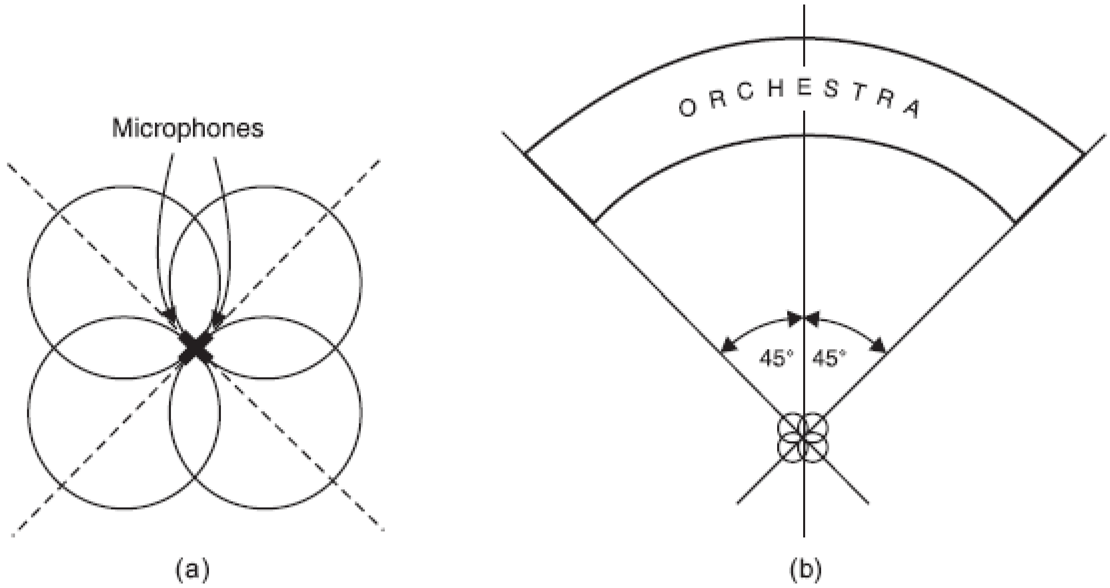


Fig. 8.5 (a) Polar characteristics of coincident microphones (b) Coincident microphones used to record an orchestra

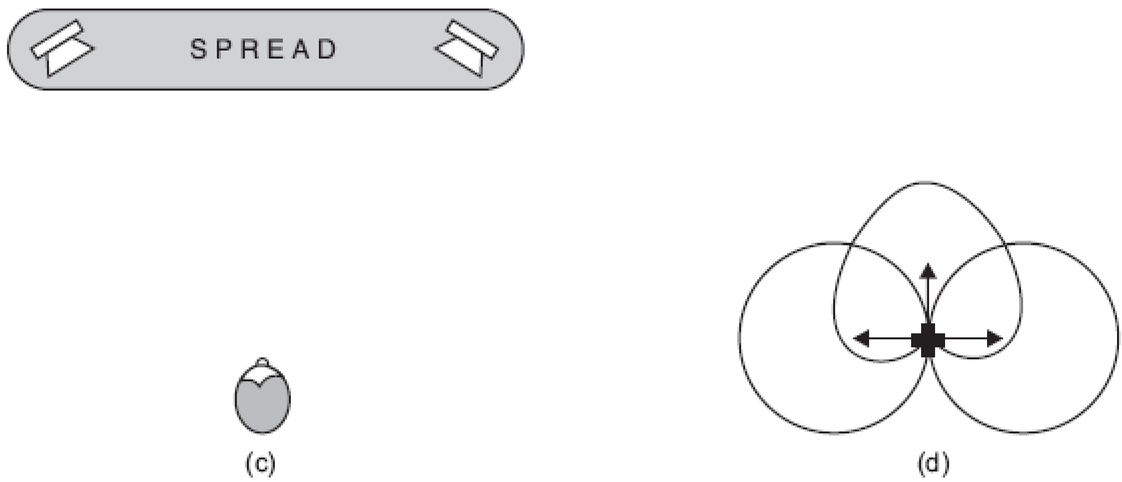


Fig. 8.5 (c) The technique shown results in a spread of sound during reproduction between speakers (d) Cardioid and bi-directional microphones are combined in the middle-side recording system

STEREO RECORDING SYSTEMS

Stereo records can be produced in two ways. In one method, one track is cut in the record groove by the side-to-side (*lateral*) movement of the record cutter, as in cutting a mono-record. The second track is cut simultaneously in the same groove but with an up-and-down *(hill-and-dale)* movement. This is shown in Fig 8.6 and 8.7. The cartridge in the record player must then be able to follow *both* the lateral and hill-and-dale movements.

In the second method, the groove is a V-shaped with a 90° angle between sides, as shown in Fig. 8.8. Both tracks are cut by a quasilateral method, one track on each side of the V. The two tracks are, therefore, 90° apart and require a cartridge that can follow both tracks at once.

The second method, known as the 45°– 45° system, is the one used for modern stereo recordings. The principles used in cutting grooves in this process are similar to both the older monophonic lateral and hill-and-dale methods. However, there is one important difference. *In modern stereo recording a single stylus (cutter) simultaneously cuts an independent information channel into each wall of the record groove.*

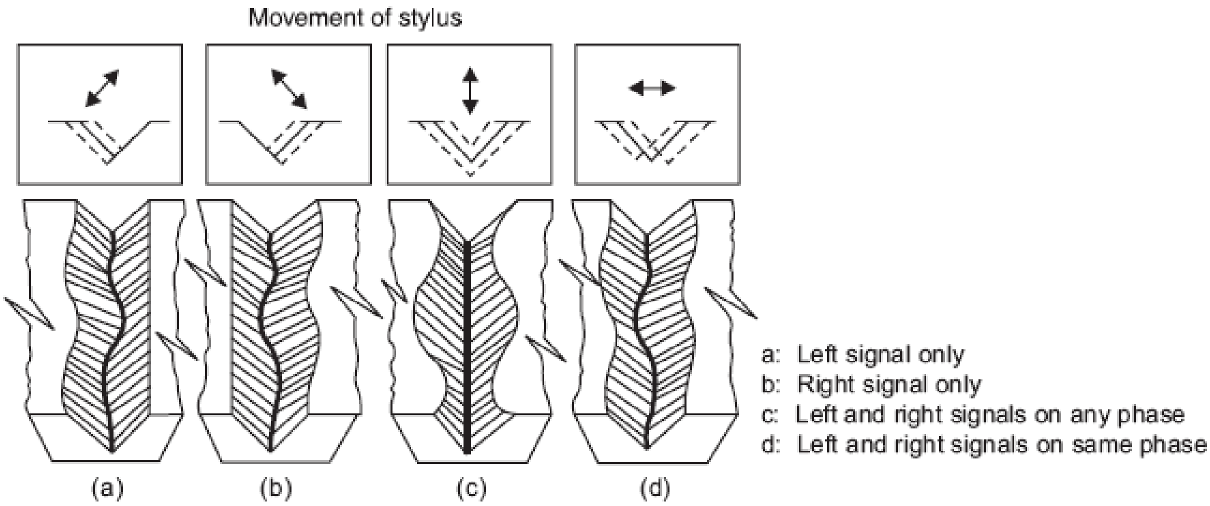


Fig. 8.6 Stylus movement in stereo-record grooves

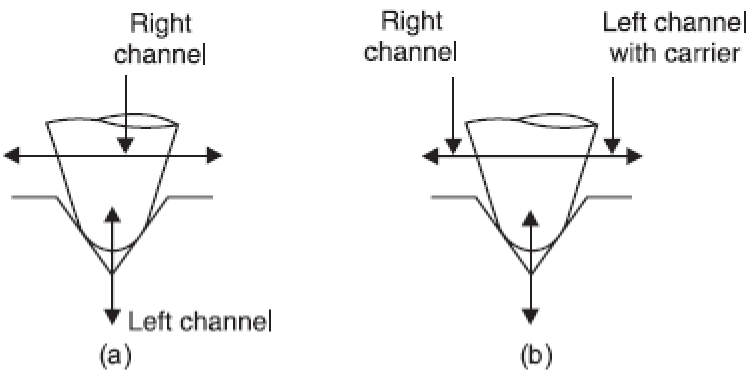


Fig. 8.7 Other systems of stereo recording (a) Vertical/lateral system (b) Frequency modulated lateral system

EXTRACTING THE MUSIC

There is only one way to extract the music and that is with a cartridge. A cartridge (Fig. 8.9) consists of three basic elements : the *stylus*, the *cantilever*, and the *generating system.*

1. **The stylus :** *The stylus is the only part which makes contact with the record.* A force must be exerted by the tonearm Fig. 8.10, to keep the stylus in the centre of the groove when the record is spinning.

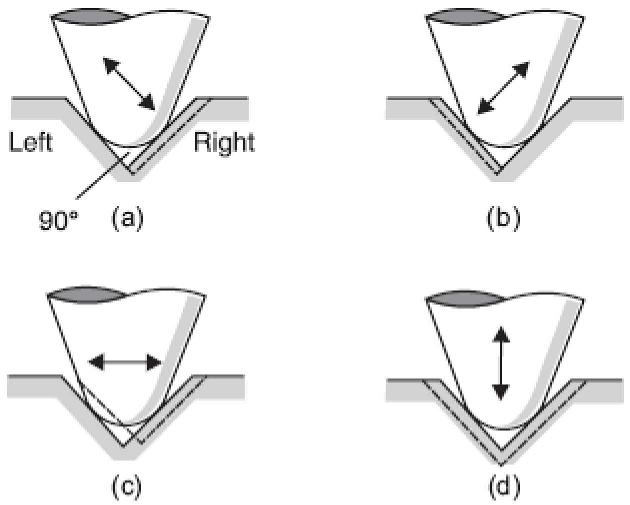


Fig. 8.8 Movement of the stylus in a stereo groove (a) Right-hand channel modulated. (b) Left-hand channel modulated. (c) Both channels modulated equally and in phase. (d) Both channels modulated equally but in opposite phase.

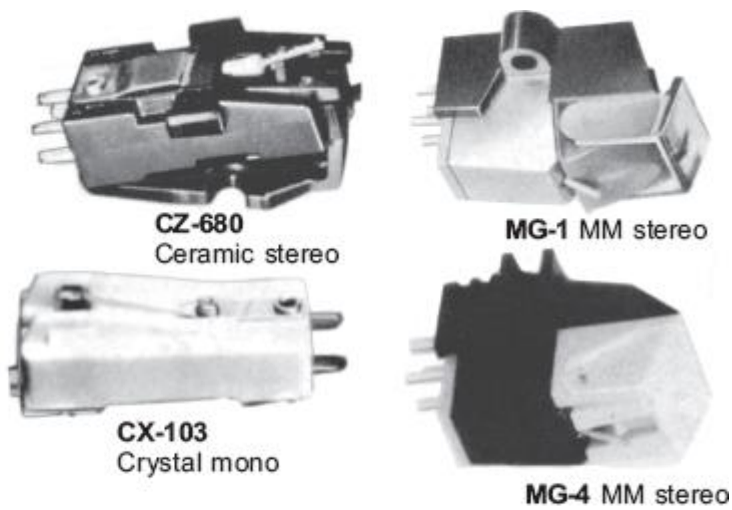


Fig. 8.9 Mono and stereo cartridges

This tracking weight is usually low—between one and two grams—but because the stylus makes contact with less than one millionth of a square inch of the record surface it exerts tremendous pressure : 6000 pounds per square inch. With such enormous pressure any roughness or irregularity in the stylus will cause record damage. The stylus must also be hard or it will wear out quickly. For these reasons, high-quality styluses are made from the hardest material known to man—*diamond.*

There are three common shapes of styluses, Fig 8.11; spherical stylus, elliptical stylus, and fine-line or line-cut stylus.

Fig. 8.10 The Tonearm

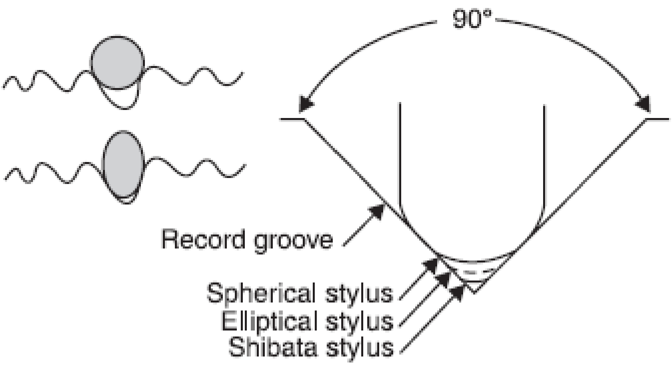


Fig. 8.11 In theory, an elliptical point reads the recorded waveform more accurately than does a conical point.

1. *Spherical Stylus:* The spherical (also called conical) shape is the least expensive to manufacture and is used in low cost cartridges. Unfortunately, *this stylus cannot accurately trace the highest musical tones.*
2. *Elliptical Stylus :* The more expensive elliptical stylus has a narrow profile enabling it to follow the complex undulations of the groove with greater precision. *It is used in the majority of high quality cartridges.*
3. *Fine-Line Stylus :* A third stylus shape, which goes under various names such as fine-line and line cut, provides more faithful tracking of the groove. It has an even narrower profile than the elliptical shape and makes contact with a wider area of the groove, thus

reducing the tremendous pressure the stylus exerts on the delicate record surface. This shape also reduces record wear.

2. **The Cantilever :** The precious diamond moves at tremendous speed through the complex groove of a spinning record. A delicate armature, called *the cantilever*, transmits its motions to the generating system, (see Fig. 8.12).

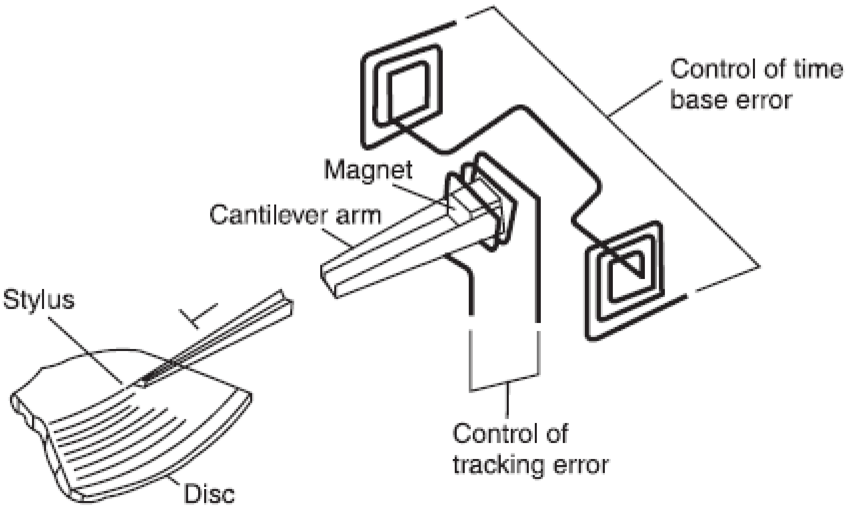


Fig. 8.12 Cantilever arm

In order for the cantilever to relay the minute movements of the stylus, it should be light weight so it can respond rapidly to the twists and turns of the record groove. And it must be rigid because any bending or flexing adds distortion and sound colouration. Obviously, light weight and high strength require great ingenuity in design and materials.

3. **The Generating System :** High fidelity cartridges are based on the law of physics which states : when a material capable of conducting electricity is set into motion in a magnetic field—or when a magnet is near a conductive material (such as copper coils)—electrical signals are generated. *The direction and speed of the movement determines the ultimate strength of the generated signal.*

Thus the generating system of a cartridge converts the movement of the stylus and cantilever into what is termed the output signal. There are three ways of doing this : moving the magnet, moving the iron and moving the coil.

CHAPTER 14

OPTICAL RECORDING AND REPRODUCTION

Compact disc (CD) players are a very specialised form of phonograph, record player, or turnable. A CD player plays pre-recorded discs (carrying music, speech, etc.) through a conventional hi-fi or stereo system (amplifier and loudspeakers). The disc is single-sided, 4.75 in (120 mm) in diameter, and can contain upto 60 min of hi-fi stereo sound. The compact disc spins at a high rate of speed compared with a conventional audio record, and uses a light beam/optical pick-up instead of the standard stylus/needle pick-up arm. In addition to superior sound can provide immediate access to audio at any part of the disc. It is also possible to program CD players to play only selected portions of the audio material.

DISC

In the *Laser Vision System*, Fig. 14.1 (a), which records *video* information, the signal is recorded on the disc in the form of a *spiral track* that consists of a succession of *pits*. The intervals between the pits are known as *lands*. The information is present in the track in *analog* form. Each *transition* from land to pit and vice versa marks a zero crossing of the modulated video signal. On the *compact disc*, Fig. 14.1 (b), the signal is recorded in a similar manner, but the information is present in the track in *digital* form. Each pit and each land represents a series of bits called *channel bits*. After each land/pit or pit/land transition there is a 1, and all the channel bits in between are 0, (see Fig. 14.2).

The *density of the information* on the compact disc is very high; the smallest unit of audio information (the *audio bit)* covers an area of 1 μm^2 on the disc, and the diameter of the *scanning light spot* is only 1 μm. The *pitch* of the track is 1.6 μm the *width* 0.6 μm and the *depth* 0.12 μm. The *minimum length* of a pit or the land between two pits is 0.9 μm, the *maximum length* is 3.3 μm. The side of the transparent carrier material T in which the pits P are impressed, the upper side during playback if the spindle is vertical, is covered with a *reflecting layer R* and a *protective layer P The track is optically scanned from below the disc at a constant velocity of 1.25 m/s.* The *speed of rotation* of the disc therefore varies from about 8 rev/s to about 3.5 rev/s (or 480 rpm to about 210 rpm).

PROCESSING OF THE AUDIO SIGNAL

For *converting* the analog signal from the microphone into a digital signal, pulse-code modulation (PCM) is used. In this system the signal is periodically sampled and *each sample is translated into a binary number*. From Nyquist's sampling theorem *the frequency of the sampling should be at least twice as high as the highest frequency to be accounted for in the analog signal*. The number of bits per sample determines the *signal-to-noise ratio* in the subsequent reproduction.

In the *compact disc system* the analog system is sampled at a rate of 44.1 kHz, which is sufficient for the reproduction of the maximum frequency of 20 kHz. The signal is quantized by the method of uniform or linear quantization, the sampled amplitude is divided into equal parts. The number of bits per sample (these are called *audio bits)* is 32 ; i.e. 16 for the *left* and 16 for the *right* audio channel. This corresponds to a *signal-to-noise ratio* of more than 90 dB. The net *bit rate* is thus $44.1 \times 10^3 \times 32$ or 1.41×10^6 audio bits per second. The audio bits are grouped into *frames*, each containing six of the original samples.

Successive blocks of audio bits have blocks of *parity bits* added to them in accordance with a coding system called *Cross-Interleaved Reed-Solomon Code* (CIRC). This makes it possible to correct *errors* during the reproduction of the signal. The ratio of the number of bits before and after this operation is 3 : 4. Each frame then has *Control and Display* (C & D) bits added to it; one of the functions of C & D bits is providing the *information for the listener*. After this operation the bits are called *data bits*.

Next, the bit stream is modulated, that is to say the data bits are translated into channel bits which are suitable for storage on the disc, (see Fig. 14.2 *b*). The *Eight-to-Fourteen Modulation* (EFM) is used for this purpose. In EFM code blocks of eight bits are translated into blocks of fourteen bits. The blocks of fourteen bits are linked by these *merging bits*. The ratio of the number of bits before and after modulation is thus 8 : 17.

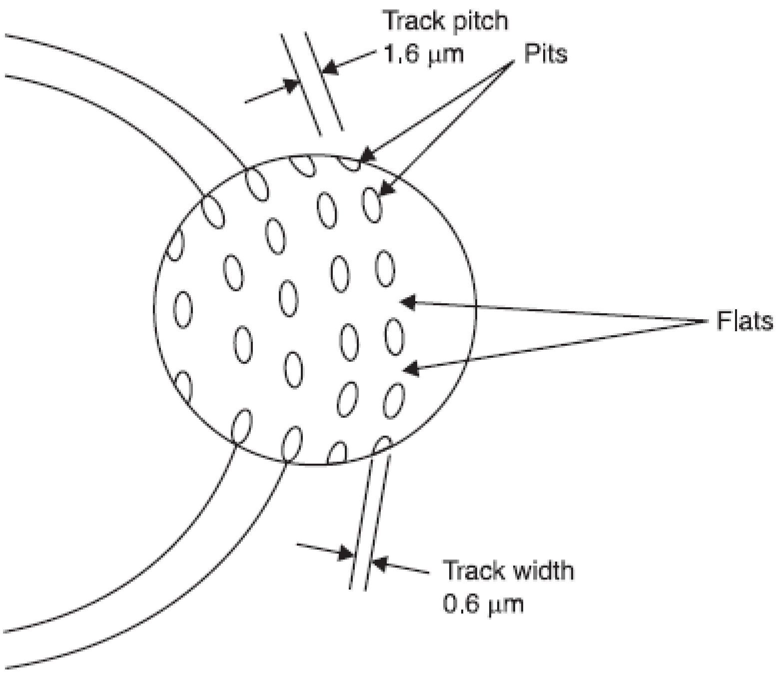


Fig. 14.3 The information on the compact disc is recorded in digital form as a spiral track consisting of a succession of pits. The pitch of the track is 1.6 μm, the width 0.6 μm and the depth of the pit 0.12 μm. The length of a pit or the land between two pits has a minimum value of 0.9 μm and a maximum value of 3.3 μm.

Table 14.1 Names of the successive signals, the associated bit rates and operations during the processing of the audio signal.

Name	**Bit rate in 10^6 bits/s**	**Operation**
Audio Signal		PCM (44.1 kHz)

Audio bit stream	1.41	CIRC + (parity bits); Addition of C & D bits
Data bit stream	1.94	EFM; Addition of merging bits
Channel bit stream	4.32	Addition of synchronisation patterns.

For the synchronisation of the bit stream an *identical synchronisation pattern* consisting of 27 channel bits is added to each frame. The *total bit rate* after all these manipulations is 4.32×10^6 channel bits/s.

Table 14.1 gives a survey of the *successive operations* with the associated bit rates, with their names. From the magnitude of the channel bit rate and scanning speed of 1.2 m/s it follows that the *length of a channel bit on the disc is approximately 0.3 µm.*

The signal produced in this way is used by the manufacturer to *switch On and Off the laser beam* that illuminates the light sensitive layer on a rotating glass disc, called the *master*. A pattern of pits is produced on the disc by means of a photographic developing process. After the surface has been coated with a thin silver layer, an electroplating process is applied to produce a nickel impression called the *metal father*. From this father disc, impressions called *mother discs* are produced in a similar manner. The impressions of the mother discs, called *sons or stampers*, are used as tools with which the pits *P* are *impressed* into the thermoplastic transparent carrier material T of the disc.

READOUT FROM THE DISC

The disc is *optically scanned* in the player. This is done by *AlGaAs semiconductor laser* Fig. 14.4 shows the optical part of the pickup. The light from the *laser* La (wavelength 800 nm) is focused through the *lenses* L_2 and L_1 onto the *reflecting layer* of the disc. The diameter of the light spot S, Fig. 14.5, is about 1 µm. When the light falls on an *interval* between two pits, the light is almost totally reflected and reaches the four photodiodes D_1 to D_4 via the half-silvered mirror M. When the spot lands on a *pit*—the depth of a pit is about 1/4 of the wavelength in the transparent substrate material—interference causes less light to be reflected and an appreciably smaller amount reaches the photodiodes. When the output signals from the four photodiodes are added together the result is a *fairly rough approximation* to the rectangular pulse pattern present on the disc in the form of *pits and intervals.*

The optical pick-up shown in Fig. 14.4 is very small (about 45 × 12 mm) and is mounted in a *pivoting arm* that enables the pick-up to describe a *radial arc* across the disc, so that it can scan the complete spiral track. Around the pivotal point of arm is mounted a *linear motor* that consists of a combination of a coil and a permanent magnet. When the coil is energised the pick-up can be directed to any required part of the track, the *locational information* being provided by the C & D bits added to each frame on the disc. The pick-up is thus able to find *independently* any particular passage of music indicated by the listener. When it has been found the pick-up must then *follow the track accurately* to within ±0.1 µm without being affected by the next or previous track. Since the track on the disc may have some slight *eccentricity*, and since also the suspension of the turntable is not perfect, the track may have a maximum *side-to-side swing* of 300 µm. A *tracking servo system* is therefore necessary to ensure that the deviation between pick-up and track is smaller than the permitted value of + 0.1 μm and in addition, to absorb the consequences of small vibrations of the player.

The *tracking-error signal* is delivered by the four photodiodes D_1 to D_4. When the spot *S*, seen in the radial direction, is situated in the centre of the track, a *symmetrical* beam is reflected. If the spot lies slightly to one side of the track, however, interference effects cause *asymmetry* in the reflected beam. This asymmetry is detected by the prism P_2, which *splits* the beam into two components. Beyond the prism, one component has a higher mean intensity than the other. The signal obtained by coupling the photodiodes as $(D_1 + D_2) – (D_3 + D_4)$ can therefore be used as a tracking error signal.

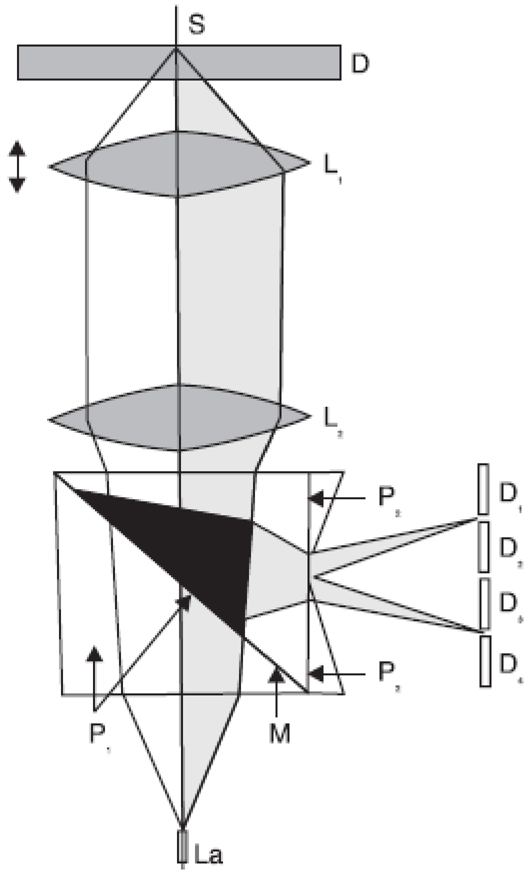


Fig. 14.4 Diagram of the optical pick-up. D, radial section through the disc. S laser spot, the image of the disc of light emitting part of the semiconductor laser La. L_1, objective lens, adjustable for focusing, L_2 lens for making the divergent laser beam parallel. M half-silvered mirror formed by a film evaporated on the dividing surface of the prism combination. P,, P_2 beam splitter prisms, D_1 to D_4 photodiodes whose output currents can be combined in various ways to provide the output signal from the pick-up and also the tracking-error signal and the focusing-error signal. (In practice the prism P_2 and the photodiodes D_1 to D_4 are rotated by 90° and the reflection of the mirror M does not take place in a radial plane but in a tangential plane).

As a result of the *aging or soiling of the optical system*, the reflected beam may acquire a slowly increasing, more or less constant asymmetry. Owing to a dc component in the tracking error signal, the spot will be *slightly off-centre* of the track. To compensate for this effect a *second* *tracking error signal* is generated. The coil that controls the pick-up arm is therefore supplied with an alternating voltage at 600 Hz, with an amplitude that corresponds to a *radial displacement* of the spot by +0.05μm. The output sum signal from the four photodiodes which is at a *maximum* when the spot is at the centre of the track is thus modulated by an alternating voltage of 600 Hz. The *amplitude* of this 600 Hz signal *increases* as the spot moves off-centre. In addition the *sign* of the 600 Hz error signal *changes* if the spot moves to the other side of the track. This second tracking-error signal is therefore used to correct the error signal mentioned earlier with a direct voltage. The output sum signal from the photodiodes, which is processed in the player to become the audio signal, is thus returned to its maximum value.

The *depth of focus* of the optical pick-up at the position S (see Fig. 14.4) is about 4 μm The *axial deviation* of the disc, owing to various mechanical effects, can be maximum of 1 mm. It is evident that a *servo system* is also necessary to give correct focusing of the pick-up on the reflecting layer. The objective lens L_1 can therefore be displaced in the direction of its optical axis by a combination of a coil and a permanent magnet, in the same way as in a loudspeaker. The *focusing-error signal* is also provided by the row of photodiodes D_1 to D_4. If the spot is *sharply focused* on the disc, two sharp images are precisely located between D_1 and D_2 and between D_3 and D_4. If the spot is *not sharply focused* on the disc, the two images on the photodiodes are not sharp either and have also moved closer together or further apart. The signal obtained by connecting the photodiodes as $(D_1 + D_4) – (D_2 + D_3)$ can therefore be used for controlling the focusing servo system. The deviation in focusing then remains limited to + 1μm.

RECONSTITUTION OF THE AUDIO SIGNAL

The signal read from the disc by the optical pick-up has to be reconstituted to form the analog audio signal. Fig. 14.6 shows the block diagram of the *signal processing* in the player.

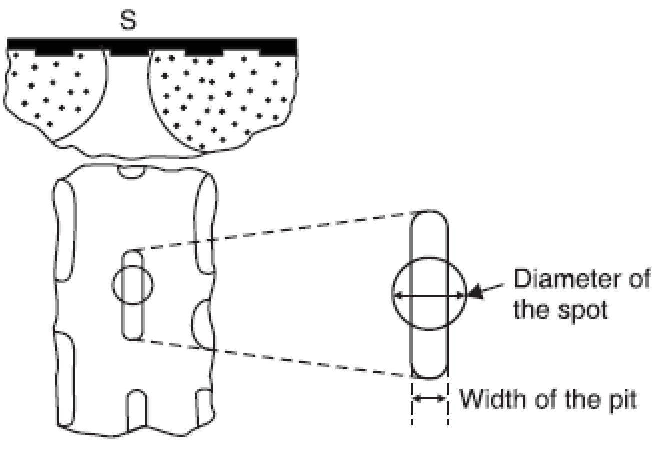


Fig. 14.5 A magnified view of the light spot S and its immediate surroundings, with a plan view. It can clearly be seen that the diameter of the spot (about 1 μm) is larger than that of the width of the pit (0.6 μm).

In Demod the *demodulation* follows the same rules that were applied to the EFM modulation, but now in the *opposite sense.* The information is then temporarily stored in a *buffer memory* and then reaches the *Error-Detection and Correction Circuit* (ERCO). The parity bits can be used here to correct errors, or just to detect errors if correction is found to be impossible. These *errors* may originate from a defect in the manufacturing process, damage during use, or finger marks or dust on the disc. Since the information with the CIRC code is *interleaved in time*, errors that occur at the input of ERCO in one frame are spread over a large number of frames during decoding in ERCO. This increases the probability that the maximum number of correctable errors per frame will not be exceeded. A flaw, such as a scratch, can often produce a train of errors called an *error burst.* The error-correction code used in ERCO can correct a burst of about 4000 data bits, largely because the errors are spread out in this way.

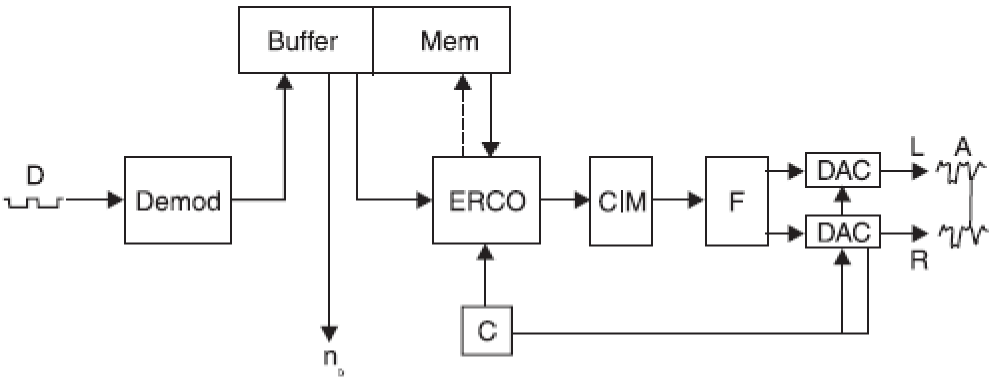


Fig. 14.6 Block diagram of the signal processing in the player. D, input signal read by the optical pick-up. A, the two output analog audio signals from the left (L) and the right (R) audio channels. Demod, demodulation circuit ERCO, error-correction circuit. Buffer, buffer memory forming part of the main memory, Mem, associated with ERCO. CIM (Concealment, Interpolation and Muting) circuit, in which errors that are only detected since they cannot be corrected are masked or concealed. F, filters for interpolation. DAC, digital-to-analog conversion circuits. Each of the blocks mentioned here are fabricated in VLSI technology. C, clock generator controlled by a quartz crystal. The degree to which buffer memory capacity is filled serves as a criterion in controlling the speed of the disc.

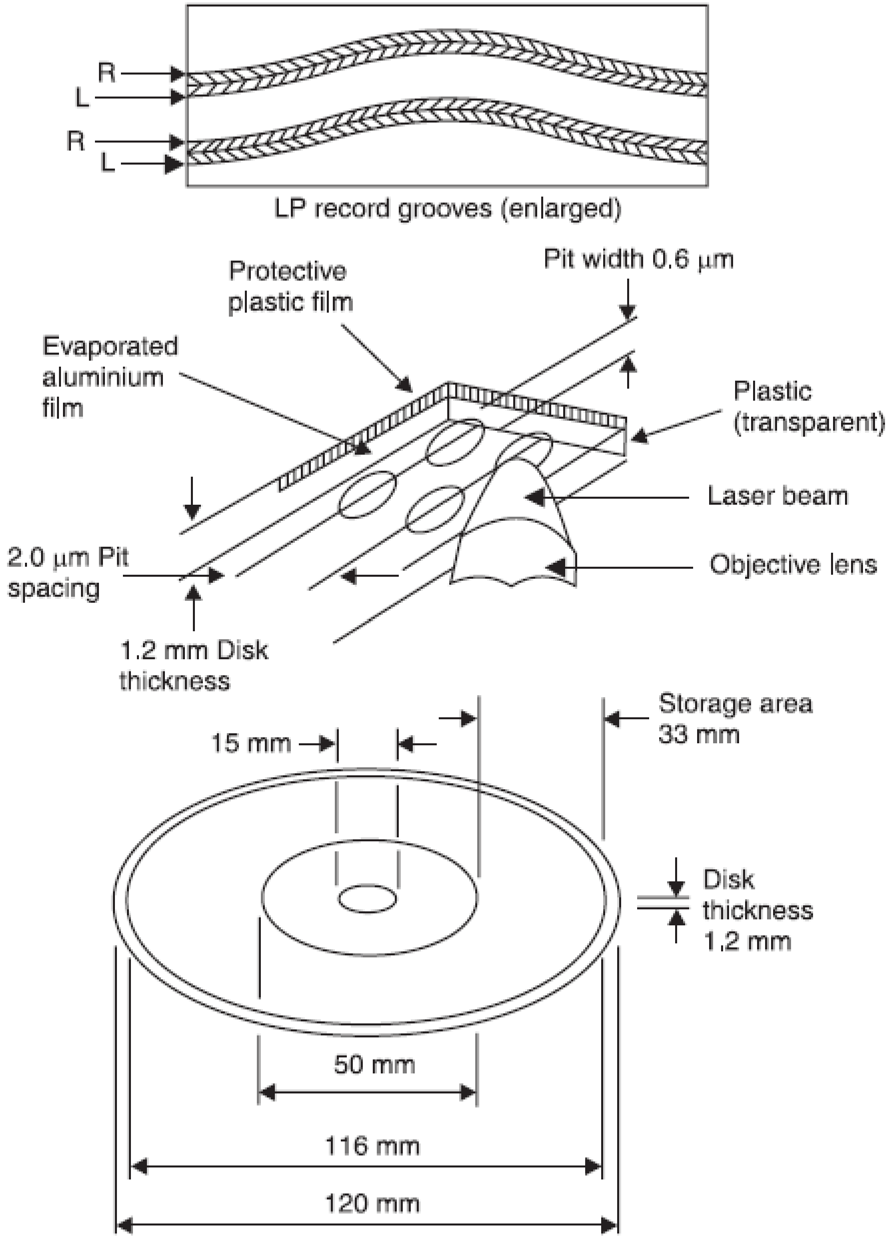


Fig. 14.7 Comparison of the grooves in a conventional LP record and the pits of a compact disc

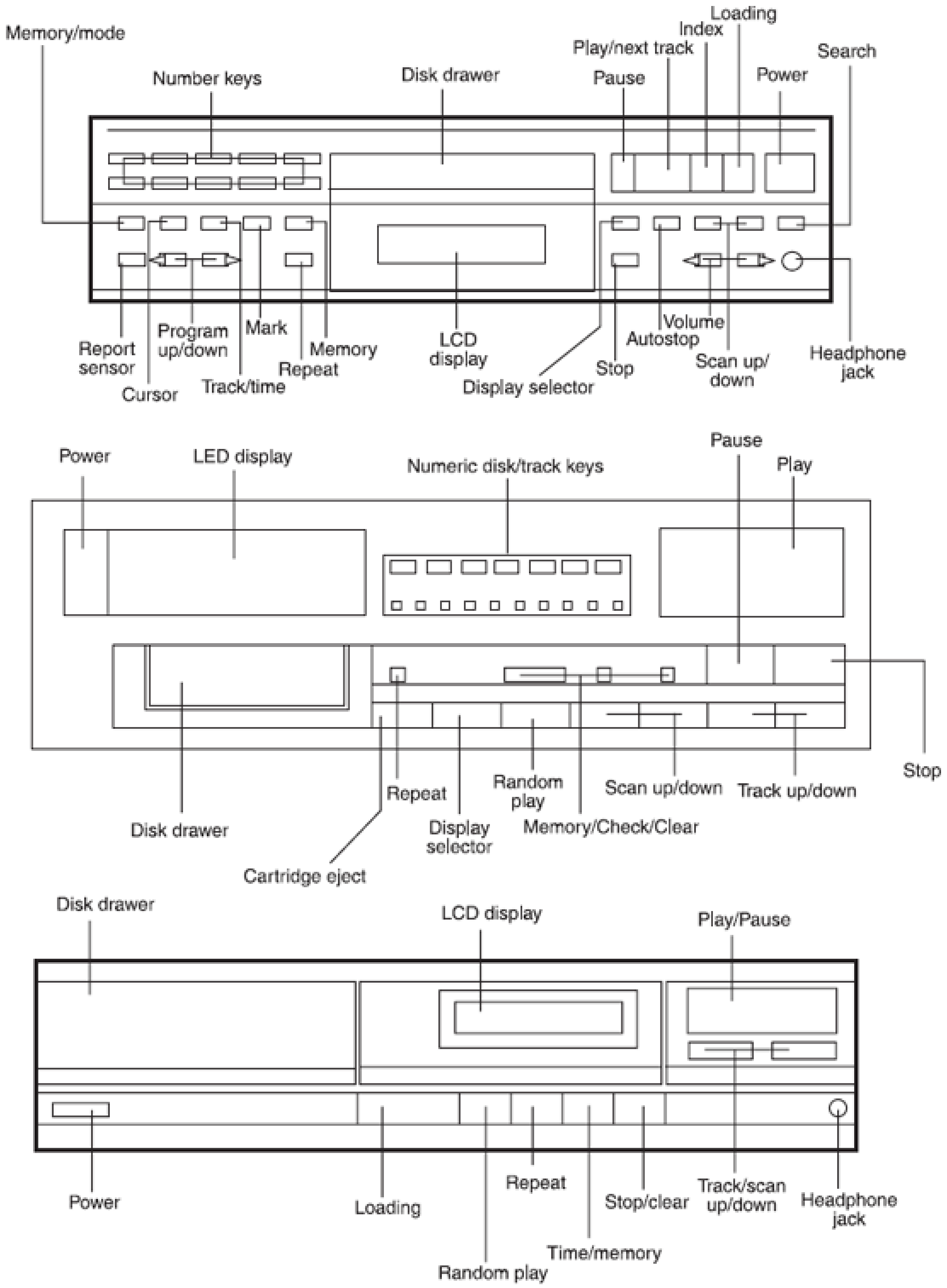


Fig. 14.8 Operating controls of typical CD players

If more errors than the permitted maximum occur, they can only be detected. In the *Concealment, Interpolation and Muting* (CIM) block, the errors detected are then masked. If the value of a sample indicates an error, a new value is found by *linear interpolation* between the preceding value and the next one. If two or more successive sample values indicate an error, they are *muted* (made equal to zero). At the same time a gradual transition is created to the values preceding and succeeding it by causing a number of values before the error and after it to decrease to zero in a particular pattern.

In the *digital-to-analog converter* (DAC) the 16 bit samples first pass through *interpolation filters F* and are then translated and recombined to re-create the original analog signal *A* from the two audio channels *L* and *R*. Since samples must be recombined at exactly the same rate as they are taken from the analog audio signal, the DACs and also CIM and ERCO are synchronised by a *clock generator, C*, controlled by a quartz crystal.

Figure 14.6 also illustrates the *control of the disc speed n*. The bit stream leaves the buffer memory at a rate synchronised by the clock generator. The bit stream enters the buffer memory, however, at a rate that depends on the speed of revolution of the disc. The extent to which n_D and the sampling rate are matched determines the *filling degree* of the buffer memory. The control is so arranged as to ensure that the buffer memory is at all times filled to 50% of its capacity. The analog signal from the player is thus completely free from *wow and flutter*, yet with only moderate requirements for the speed control of the disc.

A comparison of the grooves in a conventional LP record and the pits of a compact disc is given in Fig. 14.7.

The operating controls typical compact disc (CD) players are illustrated in Fig. 14.8.

Descriptive Questions

1. What is the difference between an LP record and a compact disc?
2. Explain the following:
    a. channel bits
    b. audio bits
    c. parity bits
    d. data bits
    e. merging bits
3. What is the function of the tracking-error signal?
4. What is the function of the focusing-error signal?
5. Explain the various steps involved in reconstitution of the audio signal.
6. Briefly explain the operating controls of a CD player.

Multiple Choice Questions

1. On the compact disc, the information is present in the track in
    a. analog form
    b. digital form
2. The density of information on a compact disc is
    a. very small
    b. very high
3. The diameter of the scanning light spot is
    a. 10 μm
    b. 1 μm
    c. 0.1 μm
4. For converting the analog signal from the microphone into a digital signal
    a. A.M. is used
    b. FM is used
    c. PCM is used

5. The track is optically scanned from
    a. above the disc
    b. below the disc
6. Locational information is provided by
    a. audio bits
    b. channel bits
    c. parity bits
    d. C & D bits

Fill in the Blanks

1. Each sample is converted into a ___________.
2. The frequency of sampling should be at least ___________ as high as the highest frequency to be accommodated in the analog signal.
3. One of the functions of C & D bits is providing the ___________ for the listener.
4. The optical pick-up is mounted in a ___________.
5. The information is temporarily stored in a ___________.
6. The extent to which n_D and the sampling rate are matched determines the ___________ of the buffer memory.
7. The information with the CIRC code is ___________ in time.
8. ___________ memory forms part of the main memory.
9. The analog signal is sampled at a rate of ___________.
10. Parity bits are added in accordance with a coding system called ___________.
11. The pick-up must follow the track accurately to within ___________.
12. A flaw, such as a scratch, can often provide a train of errors called ___________.

ANSWERS

Multiple Choice Questions

1.	(b)
2.	(b)
3.	(b)
4.	(c)
5.	(b)
6.	(d)

Fill in the Blanks

1.	binary number
2.	twice
3.	information
4.	pivoting arm
5.	buffer memory
6.	filling degree
7.	interleaved
8.	Buffer

9. 44.1 kHz
10. CIRC
11. 0.1 μm
12. error burst

CHAPTER 16

RECORDER/AMPLIFIER CIRCUITS

A 4W tape recorder amplifier and a tape preamplifier are described, and the modifications necessary for stereo operation are considered. An automatic gain control circuit for use with tape recorders is also included.

A 6W universal amplifier is described. It incorporates an equalising preamplifier, an active tone control circuit and a BEL IC-CA810 in the power output stage. The performance characteristics of the amplifier are also given.

VIDEO TAPE FORMAT PARAMETERS

Table 33.3 gives a comparison of the main physical features of each format in current use. In comparing the data, the following general points should be borne in mind.

1. Narrow video tracks give high recording density on tape but an inferior noise performance; at track widths below 25 microns good tracking is difficult to achieve with a Dynamic Track Following (DTF) or Automatic Track Finder (ATF) system.
2. Head drum diameter is directly related to video-track writing speed; the larger the drum (for a given wrap angle, i.e. 180 degree) the greater the writing speed and the wider the frequency response.
3. Linear tape speed determines the sound-channel frequency response where longitudinal tracks are used—the higher the speed the greater the frequency range.
4. Longitudinal track audio S/N ratio depends on track-width; the half-width stereo tracks for use with stationary audio heads have the worst S/N ratio.
5. All half-wrap head drums for use with 50 fields and TV systems rotate at 1500 rpm in order that one-half revolution occupies exactly the period of one TV field; in 60 Hz systems the head-drum speed is 1800 r.p.m.

Table 33.3 Physical Characteristics of the Four Main Videotape Formats

Format	VHS	Beta	V2000	Video 8
Tape	12.65	12.65	12.65/2+	8
SP linear tap speed, cm/s	2.339	1.873	2.442	2.051
Standard drum diameter, mm	62	74.5	65	40
Audio track width, mono mm	1.0	1.05	0.65	0.5*
SP video track width µm	49	32.8	22.5	34.4
Video writing speed, m/s	4.86	5.83	5.08	3.12
Video track angle to tape	5°57	5°01	2°39	4°54
Video head azimuth offset	±6°	±7°	±15°	±10°

* Only used for auxiliary purposes

+ Flip-over cassette

VIDEO TAPE TRACK CONFIGURATION

The standard VHS track pattern is shown in Fig. 33.13. The tape advances at 2.34 cm/s to give tracks 49 microns wide. They are scanned from bottom to top of the tape. The upper edge is reserved for a conventional sound track 1 mm wide, which for stereo is split into two 0.35 mm tracks separated by a 0.3 mm guard band. The lower tape edge carries the control track, which provides a poisoning reference for the video tracks themselves, and is used to guide the head sweeps during replay.

Fig. 33.14 illustrates the Betamax track pattern. Tape width is 12.65 mm again, but *video tracks are narrower* at 33 microns with a corresponding lower linear tape speed of 1.87 cm/s. The larger Beta head drum of 74.5 cm diameter is responsible for this format's *high writing speed* of 5.83 m/s compared with VHS format's 62 mm drum and 4.86 m/s writing speed.

The third format using 12.65 mm tape is the V 2000 system shown in Fig. 33.15. This is a *double-sided arrangement* with only one-half of the tape width used; in effect it is a 6.3 mm (1/4 in) system sharing with the compact audio cassette the *turnover feature* whereby both sides can be used to effectively double playing time. *V 2000 video tracks are the narrowest of all SP domestic format ones at 23.5 microns* calling for a linear tape speed of 2.442 cm/s with the specified head-drum diameter of 65 mm. In laying down these narrow tracks, and particularly in tracing them during replay, head pisitioning is governed by an *electronic steering system*, a variant of which is also used in the most recently developed format, Video 8.

Tape parameters of Video 8 are given in Fig. 33.16. *A major departure from previous practice is the tape width. This narrow tape lends iself well to portable, mobile and light weight video applications*, especially since its cassette package (95 × 62.5 × 15 mm) is a little bigger than an audio tape. Except for special applications like program indexing/cueing and other auxilary signals, the V8 format has no need of longitudinally recorded tracks. Tracking control signals in an ATF system are recorded alongwith the video signals themselves, and sound is also carried in the narrow helical tracks—*this technique of sound with vision (hi-fi) record and playback is essential in the V8 system* since the linear tape speed—approximately 2 cm/s in standard-play mode and 1 cm/s long-play mode—would render very low quality sound from a stationary-head system.

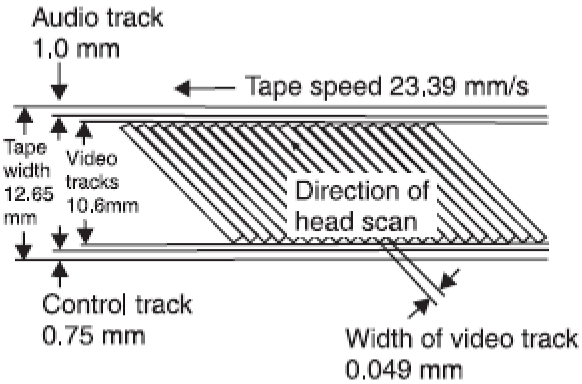


Fig. 33.13 Standard VHS track pattern

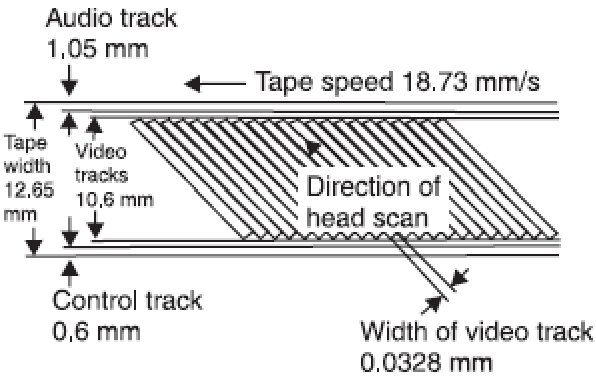


Fig. 33.14 Betamax track pattern

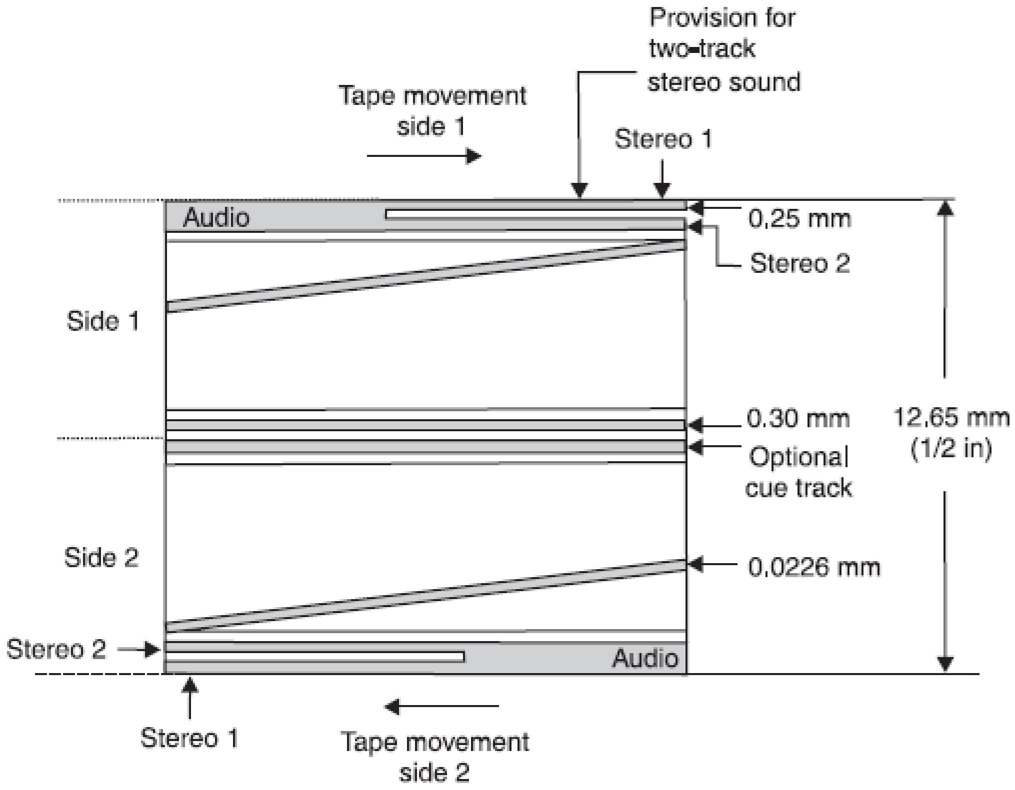


Fig. 33.15 V 2000 track pattern

Other features of V8 technology are a facility for Pulse Code Modulation (PCM) audio recording on a 30 degree extension of the video track; a relatively low writing speed of 3.1 m/s, due to the conveniently small head drum diameter of 40 mm; and a 34.5 micron track width, which in LP mode reduces to 17.2 micron.

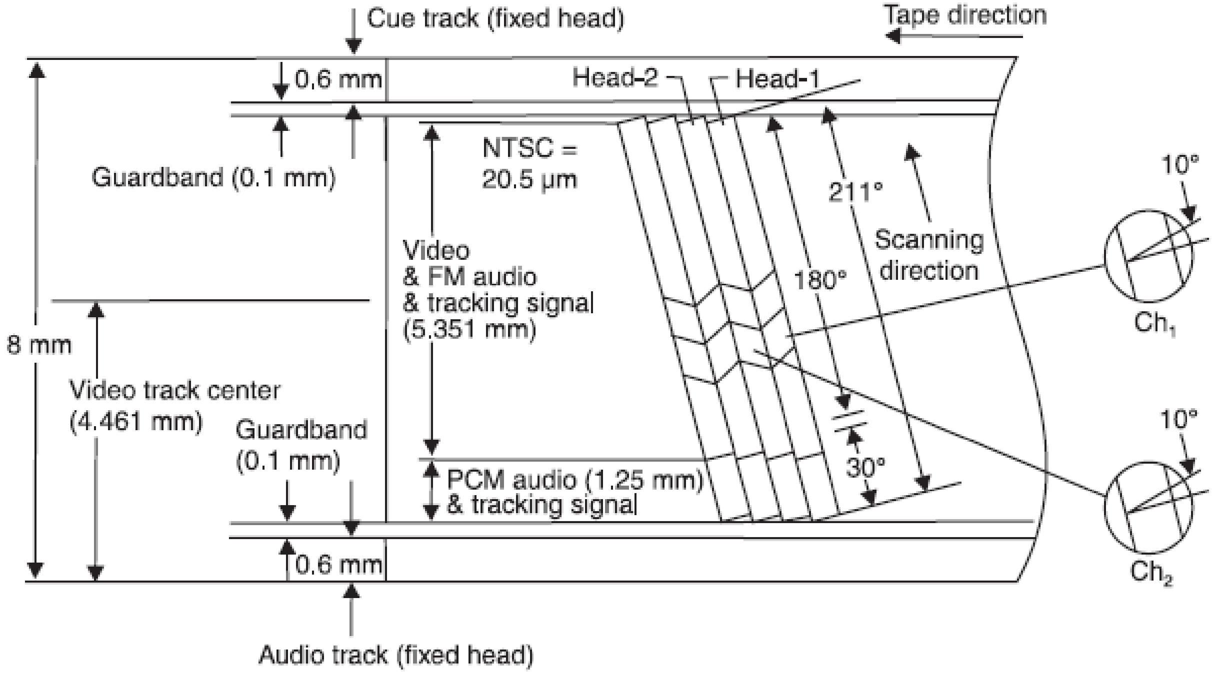


Fig. 33.16 The 8 mm track pattern

VIDEO AND AUDIO HEADS

VCRs record and playback audio and video signals using *magnetic heads.* At a maximum, all VCRs use the following magnetic heads:

1. A stationary full-erase head
2. A stationary audio head
3. Rotating video heads
4. A stationary control-track head

The *way* these heads record information on the tape is similar for the three formats, though the exact *track widths* and *specifications* vary. The recording formats for VHS, Beta, and 8 mm are shown in Fig. 33.17. *The relative location* of the heads in the three types of VCRs is shown in Fig. 33.18.

1. *Full-erase head:* The full-erase head erases the entire tape *and is on only during recording.* The erase head blanks out any previous recording that might be on the tape. In some VCRs a *secondary audio erase head* can be used. This head erases just the audio track.
2. *Audio head:* In the *record mode* the tape moves past the audio head, where the head impresses the sound signal on a thin track on the edge of the tape. In the *playback mode* the head picks up the signal recorded on tape.
3. *Video heads :* VCRs use *rotating heads* to record and playback video, Fig. 33.19. The heads are mounted in a polished metal cylinder called the *head drum* (other names are used as well, including *scanner head cylinder* and *head wheel*).

The heads are at an *angle* so they record a series of long, diagonal tracks on the tape, as shown in Fig. 33.20. The tape might *creep* slowly through the VCR, but the rotating video heads *spin* at 1500 revolutions per minute, effectively covering about 250 inches of tape per second.

Table 33.4 shows the *relative speeds* of the tape traveling through a VCR versus writing speed, which is the amount of tape covered by the video heads in one second.

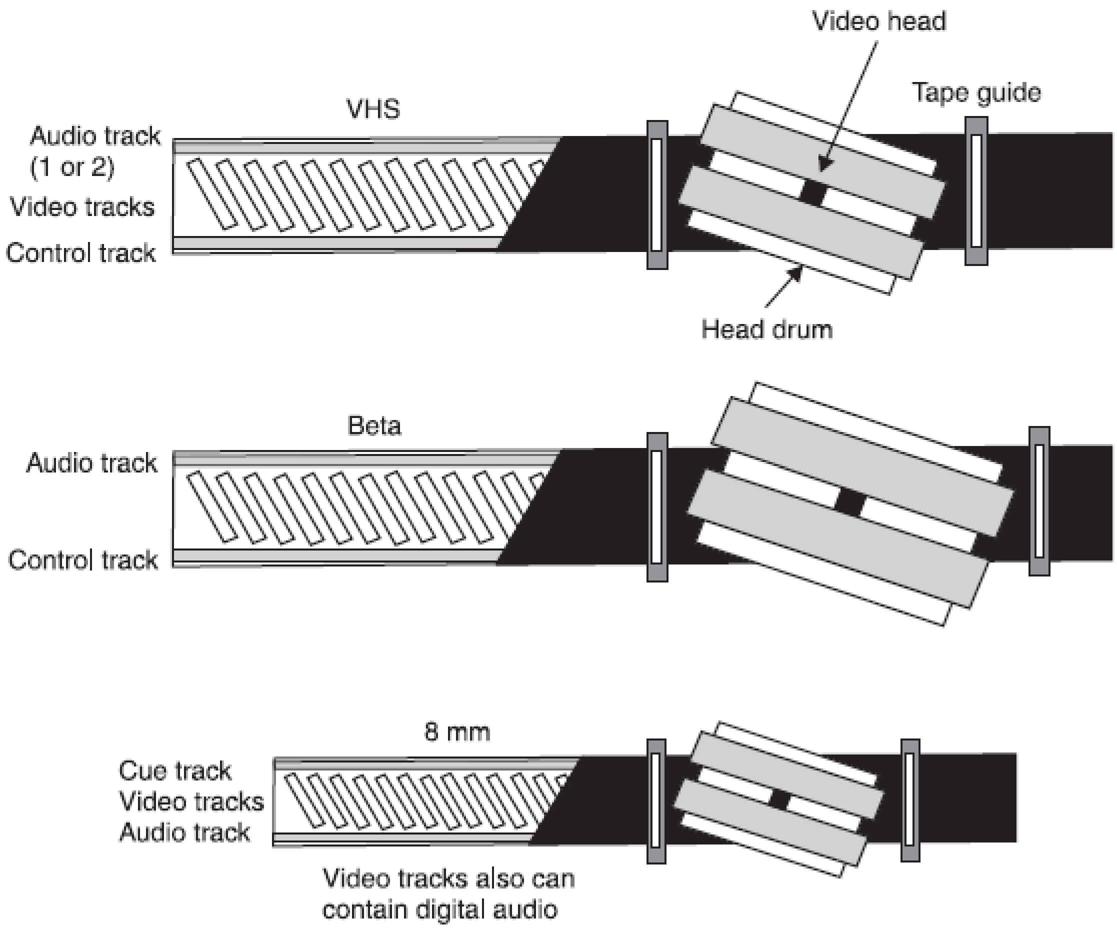


Fig. 33.17 The three home video formats and how the information tracks are recorded on each one

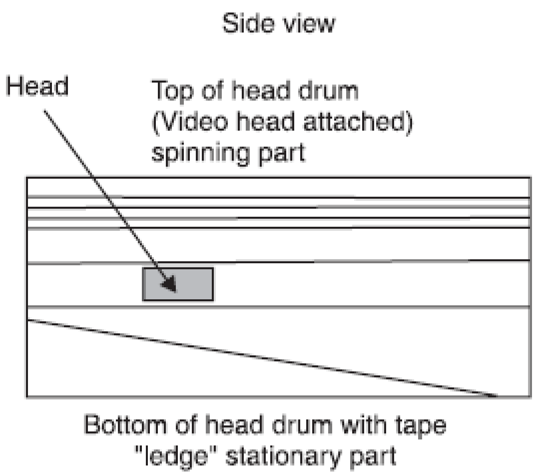


Fig. 33.18 A simplified view of a video head. The bottom of the drum is stationary; the top spins with the head.

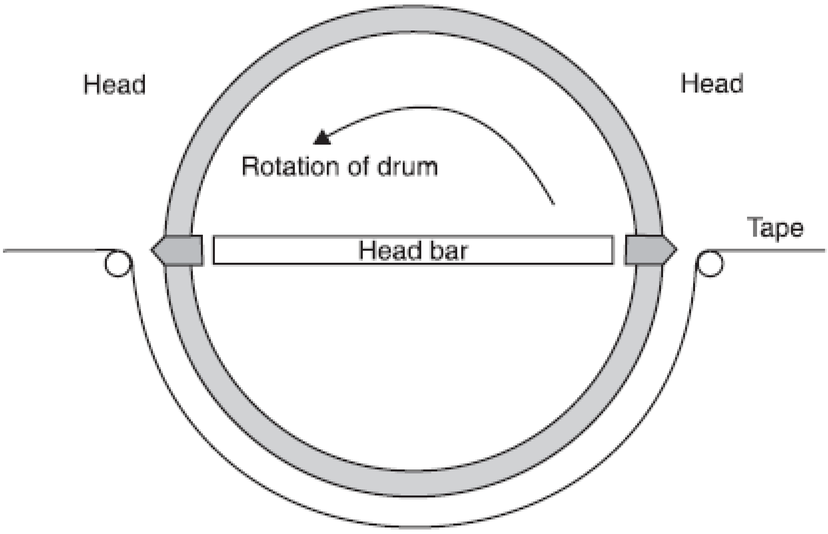


Fig. 33.19 Top of a video head drum, showing both heads and the 180 degree path of the tape.

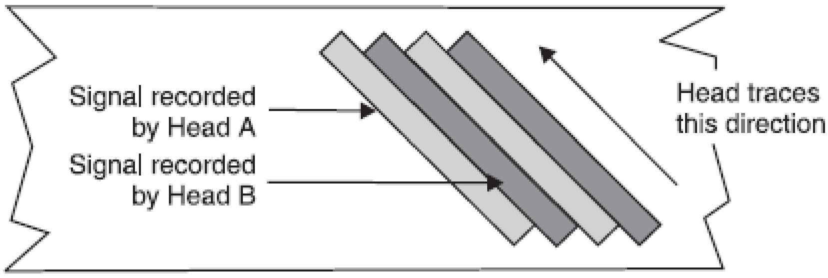


Fig. 33.20 The heads on the head drum record adjacent tracks on the tape

Table 33.4 Writing Speeds

Format and Speed	Writing Speed	Linear Tape Speed
Beta		
B-I	274.6 ips (6975 mm/s)	1.57 ips (40 mm/s)
B-II	275.4 ips (6995 mm/s)	0.78 ips (20 mm/s)
B-III	275.6 ips (7002 mm/s)	0.52 ips (13.3 mm/s)
VHS		
SP	228.5 ips (5804 mm/s)	1.31 ips (33.35 mm/s)
LP	229.1 ips (5820 mm/s)	0.66 ips (16.7 mm/s)

EP	229.3 ips (5826 mm/s)	0.44 ips (11.12 mm/s)
8 mm		
SP	147.7 ips (3751 mm/s)	0.56 ips (14.3 mm//s)
LP	148 ips (3758 mm/s)	0.28 ips (7.2 mm/s)

Note: Writing speed increases slightly at slower linear tape traveling speeds.

4. *Control-track head:* During recording, *the control track records a series of 25 Hz pulses*. These pulses are used to synchronise the video heads during playback so they pass directly over the tracks that were previously recorded. Without the control track the video heads might not scan directly over the video tracks, and the picture would be garbled. *The control tracks serve the same general purpose as sprocket holes in a movie film.* The sprocket holes help align each frame so you see a steady picture on the screen. In almost all VCRs, the control-track head is mounted in the same unit as the audio head.

CHAPTER 34

VIDEO DISC RECORDING AND PLAYBACK

The enormous popularity of the gramophone record and its playback equipment and the rapid growth of television made it an obvious idea to look at the possibility of recording video signals on a disc. The use of a disc, as an information carrier, has the advantage that one has immediate access to any program part, but of even far greater importance is its low price by using a method of production which could be similar to that of the gramophone disc.

Research on this subject had started in Philips laboratories by 1969. Earlier investigations in the same field were made, but it was only the availability of the laser as a light source with very high brightness and manufactured in an industrial way that offered the possibility of a practical realisation. It was in September, 1972 that the first demonstration for the International trade press was given with a laboratory set-up and using a master disc as the information carrier.

Subsequent demonstrations in Berlin in 1973, Tokyo in 1974 and New York in 1975 proved the feasibility of the system and created great interest. The object of these demonstrations was to prove the superiority of the optical system over the alternative systems and thus promote the creation of one standard.

OPTICAL RECORDING MEDIUMS

The storage of information by optical methods has many advantages over the conventional method of magnetic recording. As optical recording had obviously much to offer, an intense research began for materials on which information could be stored with the aid of a laser. Three classes of materials that seem suitable for the optical recording of information are: tellurium–selenium alloys, organic compounds and magneto-optical materials.

TELLURIUM–SELENIUM ALLOYS

One of the new materials for the storage of information is a *polycrystalline tellurium-selenium alloy to which small quantities of other alloys (arsenic, for example) are added to give better control of the melting point and the stability of the material.* A thin layer of the alloy is applied to a substrate. A narrow laser beam is used to melt this material locally so that holes are created with the same depth as layer. *During the readout process, with a less intense laser beam, the presence or absence of holes produces differences in the reflection of laser light. These differences in reflection represent the information in coded form.*

The disc with a tellurium–selenium alloy is ideally suitable for use as a *storage medium* for both digital data (alpha-numeric information or digital audio) and video recording.

The use of tellurium–selenium alloys also makes it possible to *record* information on a disc, *erase* it and then use the disc again to *record new information.* By choosing the energy output of the laser appropriately (compared with the level necessary for the 'hole' disc) the polycrystalline material is *melted* locally, but no holes are formed. After the laser pulse, the molten areas cool down so quickly that they solidify in a metastable amorphous phase. These *amorphous domains* reflect differently from the crystalline surroundings on readout. *Erasure* takes place when a laser with sufficiently high energy level transforms the amorphous domains in the crystalline phase.

In most applications, the disc can be *used and erased many times.* In principal, storage of both digital data and video recording is possible because of the high S/N ratio.

ORGANIC COMPOUNDS

Organic dyes exist that absorb a great deal of light and *have a high reflectance even when applied in thin layers.* These thin layers of organic compounds seem to be a promising alternative to tellurium–selenium alloys. The *memory effect* is again obtained by melting the material locally with a laser to create small *pits.* The difference from the tellurium–selenium alloy is that *these pits do not normally penetrate through to the substrate. The reflectance varies with the depth of the pit. The difference in reflection created by the pattern of pits is used when the information is being read. The melting process is irreversible, so the disc can only be written once.*

It has been found that these organic compounds retain the information just as well as the hole discs with tellurium–selenium alloys. These compounds have also been found to be highly resistant to heat and moisture. Another attractive feature is the simple *spin-coating process* for applying organic compound to the disc.

This type of disc has many applications. *The S/N ratio obtained experimentally is high enough for the storage of both digital and video information.*

EVOLUTION OF VIDEO DISC

The concept of a gramophone record that produces pictures first became a reality in the 1920s, long before the development of the first video cassette. John Logie Baird successfully created a mechanical television system by recording simple visual signals on an ordinary 78 r.p.m. gramophone record and in 1935 his invention went on sales in Selfridges. Purchasers of Baird's video disc were able to view various British public figures in profile. The invention failed to capture the imagination of the public and was eventually withdrawn.

It was 35 years before the first video disc appeared but it was never actually launched on the British market. It was supposed to be marketed by a company called Teldec, jointly by the German company *Telefunken* and *Decca* in the U.K. Made of plastic and with a playing time of only ten minutes the disc was named *TeD;* however due to a mixture of technical problems and disagreement between Teldec partners, it never became commercially available. The Teldec debacle, however, set the video disc ball rolling, prompting a burst of activity from the electronic companies around the world as they attempted to improve upon the original idea.

Philips was the first off the mark unveiling its video disc at a press conference in Eindhoven in 1972. It has taken the company another ten years to introduce the first commercially available video disc and video disc player in the U.K.

In the United States, RCA introduced its *Select Vision system*, and in Japan, government pressure for technical standardisation finally produced some results with JVC developing its *VHD video disc system* (VHD is the short form of *Video High Density).*

The use of disc, as an information carrier, has the advantage that one has immediate access to any program part, but of even greater importance is its *low price* by using a method of production which could be similar to that of the gramophone disc.

It was only the availability of laser as a light source with very high brightness and manufactured in an industrial way that offered the possibility of a practical realisation.

Subsequent demonstrations in Berlin in 1973, Tokyo in 1974, and New York in 1975 proved the feasibility of the system. The object of the demonstrations was to prove the *superiority* of the optical system over the alternative systems and thus create the promotion of one standard.

Fig. 34.1 A picture produced on a receiver by one of Baird's video discs in 1938. It lasted several seconds.

In accordance with this endeavour, an arrangement was reached with Pioneer in 1974 to combine efforts in further developing the system. At the same time there was the numerous advantage of the availability of a wealth of software for the Pioneer and Philips video disc systems, since known as *Pioneer Disco-Vision system* and *Philips VLP system* (Video Long Play). The Pioneer and Philips video disc system is much more than just an entertainment, item. *It is a new communication system offering the consumer information of his own choice at a time that suits him.* Pioneer Laser Disc Player is shown in Fig. 34.2.

Fig. 34.2 Pioneer laser disc player

The superiority of the system originates from the application of optics for reading out the information on the disc. Due to the absence of a physical contact between the disc and reading out device, *no wear occurs.* There is an effective *protection of information* on the disc against finger prints, dust and scratches. In this respect the video disc is superior to the gramophone disc! Finally, the optical system offers a number of *playing modes* to be utilised that are inconceivable with mechanical systems.

VIDEO DISC

The Pioneer and Philips video disc resembles very much in its physical shape of a gramophone records, Fig. 24.3. It consists of a transparent plastic material with *standardised diameters* of 30 cm and 20 cm and a *thickness* of 1.1 mm. The most striking differences with an audio record are *its mirror like appearance due to a reflective coating* and the fact that *it can be played on one side only.* The basic difference however, lies in the *structure of information.* Whereas the audio disc has grooves, the walls of which are modulated with an audio signal, the video disc, in view of the requirements of a much higher information density, has *tracks with a much finer information* and with a *spacing that is about 60 times smaller than that of the audio disc.*

Two basic types of disc are available; one rotating at 1800 r.p.m. for areas with NTSC television system and another rotating at 1500 r.p.m. for areas with PAL or SECAM television system. For obvious reasons, the two types are not interchangeable.

Fig. 34.3 The video disc resembles a gramophone record but has mirrorslike appearance. The player is only slightly larger than a normal hi-fi turntable.

The information on the disc exists as a *spiral track* starting at the inside of a fixed diameter and moving to the outside. Average *track pitch* is 1.6 µm, resulting in a playing time of approximately 30 minutes for a 30 cm diameter disc. The information in the track consists of small depressions called *pits* with variable distance and length. *Pit width is* 0.4 µm, where as *pit depth* is approximately 0.1 µm. Instead of depressions, other elements suitable for absorption or dispersion of light can also be used and do exist mainly as a result of the mastering system used.

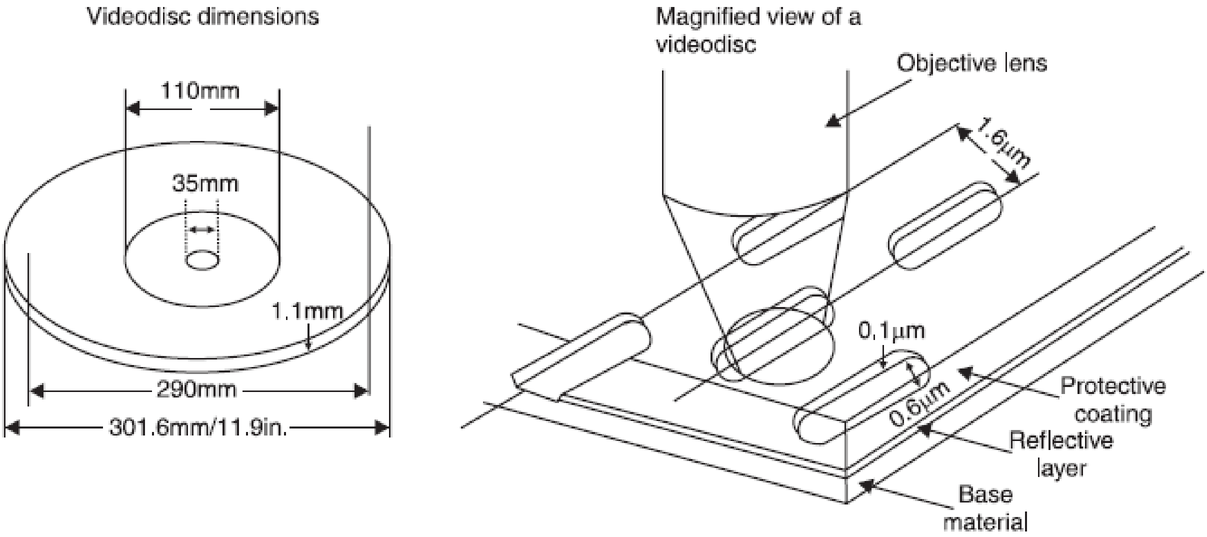


Fig. 34.4 Video disc dimensions

The *total track length* on a 34 cm disc is about 34 km and the *linear speed of tracking* varies from 37 km/h at the innermost track to 100 km/h at the outermost track. One complete television picture (two fields) requires a *surface area* on the disc of 0.5 mm at the innermost track.

VIDEO DISC MASTERING AND REPLICATION

The production process of a video disc is more or less comparable with that used for conventional gramophone record. First a *master recording* is made. It consists of a glass plate with a photosensitive layer deposited on one side. The coded signal of the information to be stored *modulates* the beam of a 1 mm laser which writes the information in the surface of the disc. Cutting is done on *real time basis;* that is it requires only as much as the program lasts and recording takes place at the disc's *rotational speed* of 30 r.p.s. for (revolutions per second) NTSC and 25 r.p.s. respectively for PAL and SECAM. In principle, every normal type of TV signal source can be connected to the cutting device. *In practice however, 50 mm magnetic recording tape is used as a program carrier.*

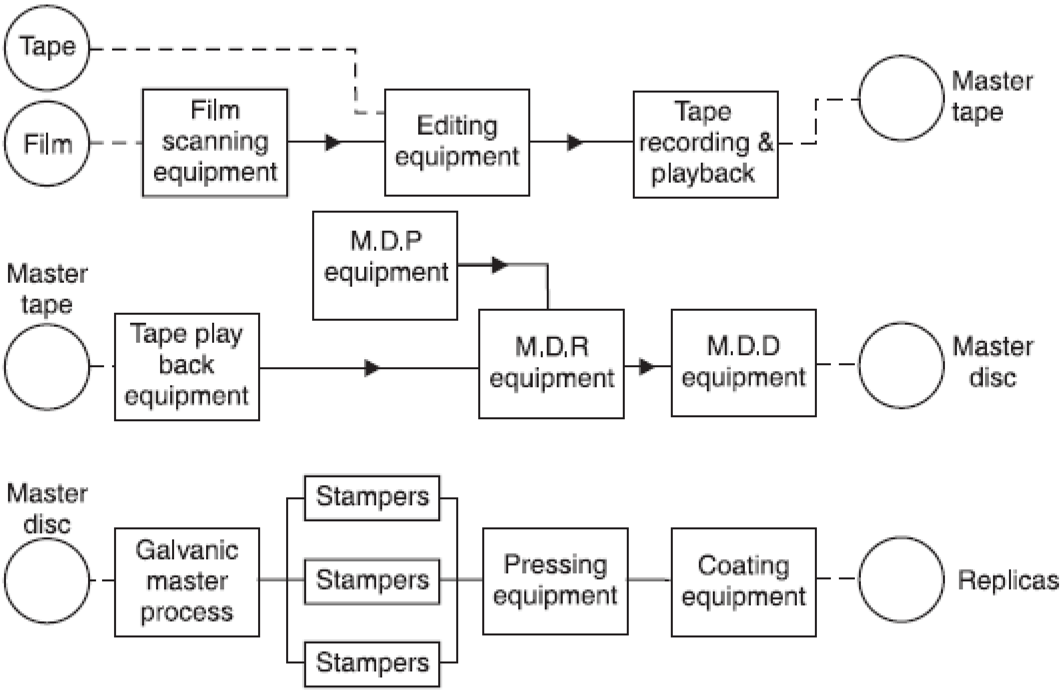


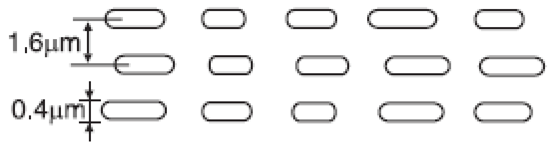
Fig. 34.5 Video disc mastering and replication

Exposure to the laser beam is followed by a development process which leaves a pattern of pits on the master from which, via a galvanic process, *stampers* are made which are used for disc production in a way similar to processing of gramophone records.

After processing, an extremely thin metal coating, not more that 0.04 μm thick is deposited on the information side which is then sealed with a *protective layer*, as shown in Fig. 34.6.

VIDEO DISC FORMATS

The three *non-interchangeable* video disc formats fall into two basic categories; optical and capacitance. The *laser optical system* (also called VLP) which is employed by Philips, uses a laser beam to recover the electronically encoded information stored on the disc. *The capacitance system (also called capacitance electronic disc or CED), employed by both JVC and RCA*, uses a stylus and tracking arm similar to that of a conventional record player to recover the information recorded on the grooves of the disc. There are two variations of the laser optical system *reflective* and *transmissive.* There are also two variations of the capacitance system, the *video/audio high density system* (VHD) and the capacitance electronic disc system (CED).



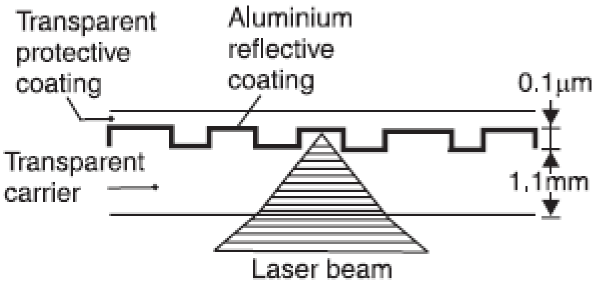


Fig. 34.6 The upper half of this diagram shows the tracks of pits, while the lower half shows a cross-section of the disc indicating how the information pits are protected by a transparent protective coating.

LASER VISION

The core of the Philips laser vision are the *shiny silver video discs* which hold the pre-recorded program material. These are *optically read* by a high precision laser, built into the laservision player. *The laservision system features a level of picture and sound without a parallel in video.* Flawless picture reproduction, full of rich colours, is accompanied by high stereo sound when the player is connected to a stereo TV or hi-fi audio system.

There are two types of Laservision discs.

Long play discs for straight forward uninterrupted playback of entertainment programs and *active play discs* which involve the viewer in learning a new skill or subject and making full use of the player's many and versatile *picture search facilities* that help you to quickly locate any point on the disc and study parts in detail.

The information etched on the *reflective material* of the disc in a *series of pits* is read off with a fine beam created by a helium–neon *laser*. By means of a series of lenses, gratings, prisms and mirrors, the laser beam is directed to the *disc underside* where it is moved by a scanning lens. When it hits a pit in the silvery surface, Fig. 34.7, the beam is reflected. The reflected beam then passes back to the photodiode which creates the signal output.

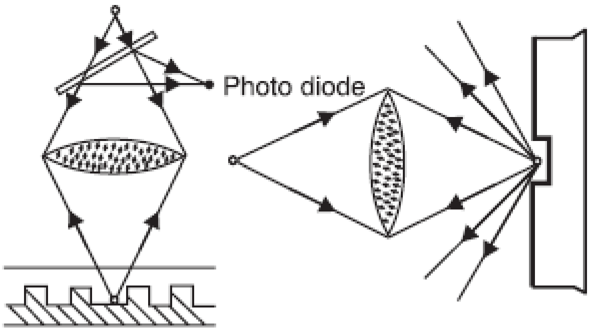


Fig. 34.7 When the scanning beam hits a pit in the silvery disc underside the beam is reflected. The beam then passes back to the photodiode which creates the signal output.

In a similar system, which is worked in France by Thomson CSF, the laser beam is not reflected from the surface of the disc, but is actually *transmitted* through it. *Variations in the quality of the beam, created by the information etched on the disc, are then used to construct the electrical signals need to create a TV picture.*

Any Laservision disc may be freely handled, cleaned and replayed as often as you wish. The program material is held in billions of tiny pits punched into the disc's surface and then covered by a transparent material which protects it against dust scratches and finger marks. Because the disc is scanned by a laser light beam, *its life is virtually infinite.* Laservision discs are marketed by a steadily increasing number of leading program distributors under their own *labels.*

SELECTAVISION (RCA)

The *cheapest and the most basic* of the three video disc formats, with the current models offering only mono sound and picture search, selectavision discs, based on the capacitance system, have a maximum *running time* of one hour per side.

The RCA stylus recovers information from the disc surface by *direct electrical means.* The selectavision disc is either formed from or coated with electrically conductive material and the stylus serves as a conductive electrode. Although the groove appears to be smooth, it does have *tiny pits along the bottom* which produce changes in electrical capacitance between the disc surface and the stylus electrode. These *changes in capacitance* are sensed by a tuned circuit to produce an output of video signals.

The TeD and RCA systems, rely on a stylus following a minute surface groove and *fail to function if the disc surface is damaged.* For this reason, the floppy TeD disc is always stored in a sleeve which is itself loaded into the player for automatic extraction of the vital foil, the sleeve remaining inside the player until the disc has played. RCA discs, which must also be warp free and are rendered useless by finger marks, which confuse the capacitance effect, are stored in a *caddy* from which the player automatically extracts the disc for playing.

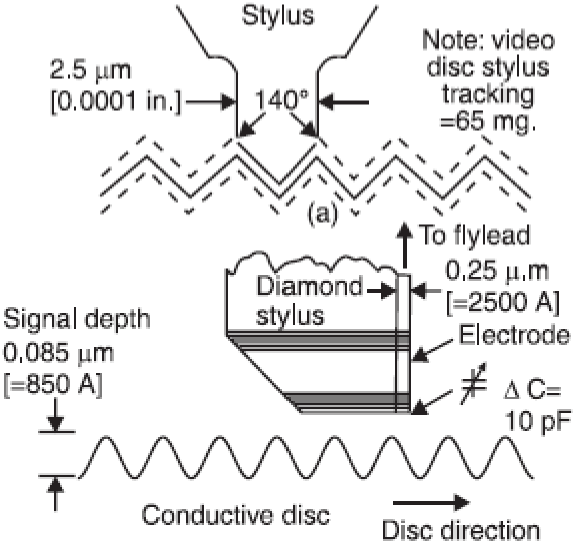


Fig. 34.8 The RCA Selectavision system employs a capacitance sensing stylus electrode, which is guided along shallow grooves and detects capacitance changes as it passes over shallow pits (a) Front view (b) Side view

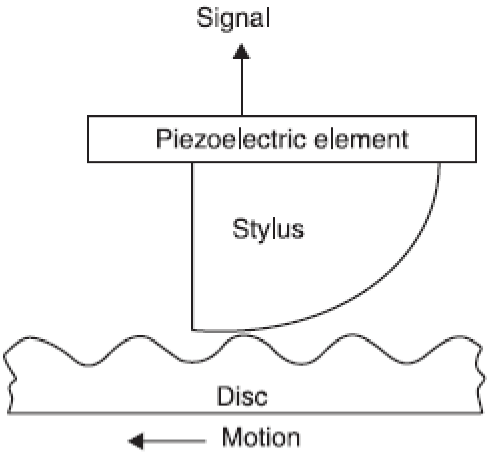


Fig. 34.9 The TeD employs a stylus shaped like the prow of a boat which senses mechanically the hill-and-dale modulations that contain the information.

VIDEO HIGH DENSITY (JVC/THORN-EMJ)

Technically and price wise, VHD falls between laservision and selectavision. The JVC system resembles the RCA approach in its use of *capacitance pick-up* from an electrically conductive disc with a spiral of pits on its surface. But the 25cm (10 inch) JVC disc has a *smooth surface with no groove.* The stylus electrode is constrained to follow the spiral pits by a servo system. The spiral of *program pits* (which the electrode stylus tracks to produce pictures on TV screen) is *interlaced* with a spiral of *tracking pits* which are sensed to control the servo system to guide the electrode.

Fig. 34.10 illustrates the method of controlling the stylus. The stylus is mounted at the end of a cantilever pick-up arm opposite that end on which a magnet is attached. Fixed coils are positioned near the magnet: a single coil is wound around but not in contact with the magnet, and a pair of vertical coils are positioned, one each on either side of the coil, and in phase opposition to each other. Thus *the stylus can move transversely and longitudinally in response to the particular current flowing in these coils. The current is varied by a tracking error signal*, or by a command to move the stylus to a desired track, permitting various functions during playback.

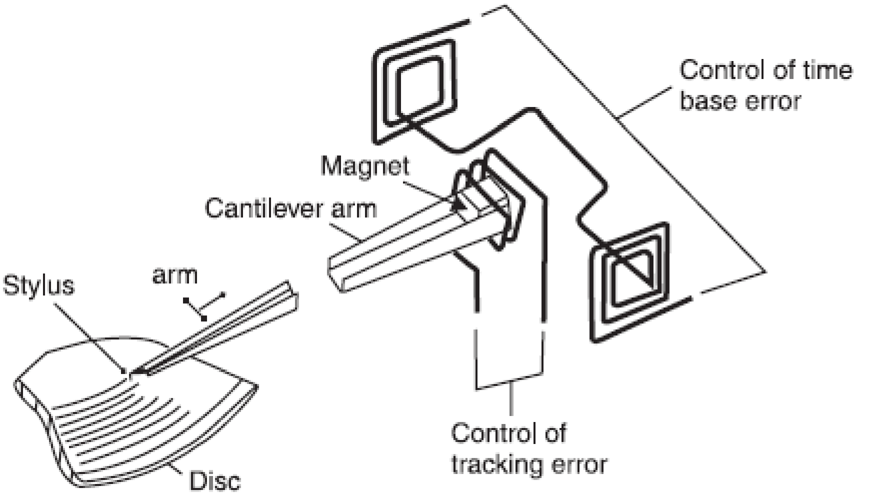


Fig. 34.10 Video signal pick-up in VHD system uses a cantilever arm

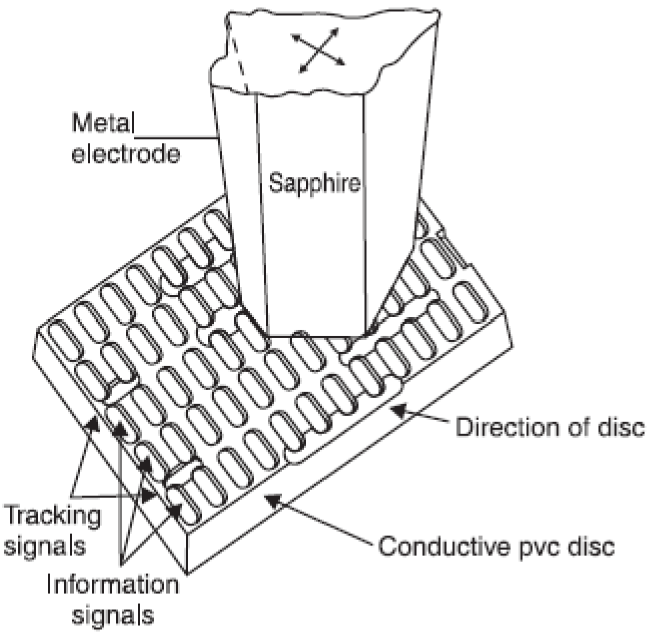


Fig. 34.11 The JVC system also employs a capacitance sensing stylus electrode. It is guided along the correct path by servo pits.

VIDEO DISC SYSTEMS — A COMPARISON

Common to all video disc systems is the process in which a *program* (originally recorded on magnetic tape) is recorded on to a *master metal disc*. The metal master is then used to mass produce *plastic discs* which are played on the video dics player (VDP).

Plastic discs for the *laser video (LV) optical pick-up system* are coated with metal on one side (the recorded surface) and then *bonded* with the metal inside for protection. Carbon is added to discs for the two *capacitive pick-up systems.* CED and VHD, to make the disc conductive. A lubricant for smoothing the pickup and reducing wear is added to the CED discs. This *lubricant* is necessary because the CED system has grooves on the disc for stylus tracking. This makes the CED system simpler (no servo tracking is required) but does produce some wear. However, the wear on a CED disc in no way compares with the wear of a conventional audio record. RCA demonstrated this by playing a single groove of a CED disc 9000 times without noticeable deterioration in the video display.

The three video disc systems also have similarities. *All three systems use a plastic disc rotating on a turntable. In all the systems, the player picks up information represented by changes in the disc surface and converts the information into signals for playback on a television set. All systems use frequency modulation (FM) for both video and audio signals. Each disk has a spiral track to carry the information rather than a series of circular tracks.*

In spite of all the basic similarities, *the systems differ not only in the pick-up technique (optical versus capacitive) but also in the format in which the information is encoded and in the method by which information is tracked. Other differences include size, material, rotation speed and signal-protection schemes.*

RECORDING SYSTEM

In the optical video disc there is a *single information track* in which all the information is stored for the reproduction of a colour television program with two sound channels and data signals.

The *nonlinearity* of the master recording process limits the choice of possible encoding techniques and a *two-level signal recording* was found to be the most attractive solution. On this track *the information* is enclosed in the length and the spacing of the pits or, in other words, for a rotating disc in the repitition frequency, determined by the average length of the pits, and a pulse width modulation of the frequency, determined by the modulation of the length of the pits, Fig. 34.12.

The *composite video signal* employed in the video disc system is frequency modulated on a carrier at 8 sMHz which is pulse width modulated by two hi-fi audio channels at 2.3 MHz and 2.8 MHz.

Fig. 34.13 shows the block diagram of the signal processing for coding the video and audio system. Before FM modulation of the *video signal* pre-emphasis time constants of 50 μs and 12.5 μs are employed. The *audio signals* are FM modulated on carriers of 2.3 MHz and 2.8 MHz with a frequency deviation of + 100 kHz and a pre-emphasis of 75 μs. The two audio carriers are summed with the FM carrier and after limiting, the output signal is used to modulate the intensity of a laser beam passing through an electro-optical modulator in the master recording machine.

The *spectrum* of video and audio signals is given in Fig. 34.14. The *master* is used as a starting point for the production of discs.

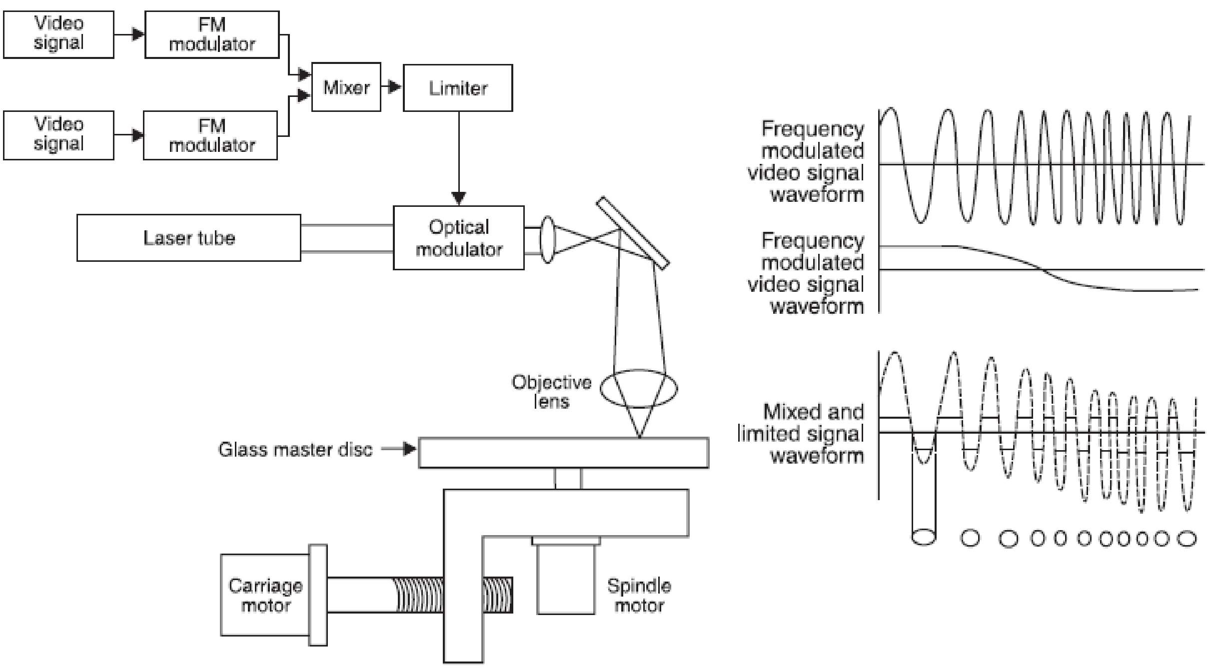


Fig. 34.12 Video disc recording system

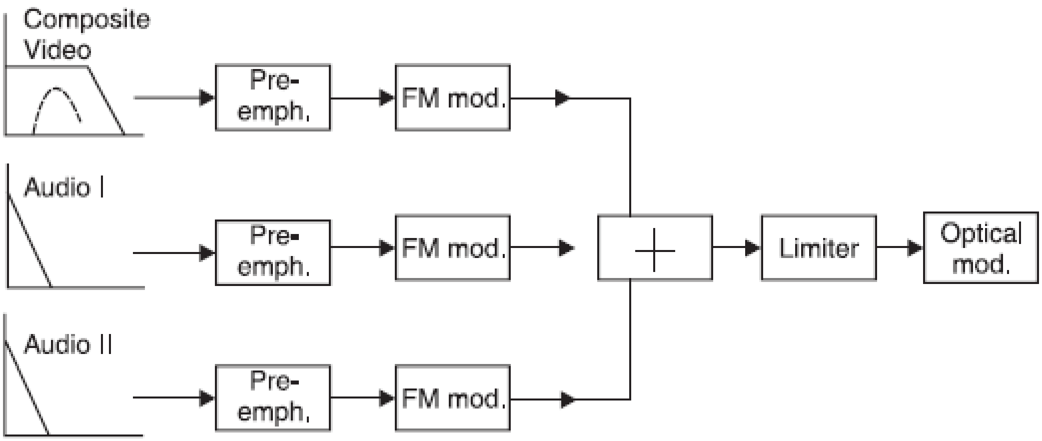


Fig. 34.13 Signal processing — encoding

PLAYBACK SYSTEM

In reading back the information, the reflected light returning from the disc falls on a photodiode and its output is *amplified and corrected* according to the frequency characteristic of the player. A high-pass filter separates the video information and the filters have a crossover frequency at 3.5 MHz. The separated FM signals are then demodulated and a *de-emphasis* is applied to compensate for the *pre-emphasis* employed before recording, in order to achieve a better S/N ratio and a more uniform frequency response. The playback system is shown in Fig. 34.15.

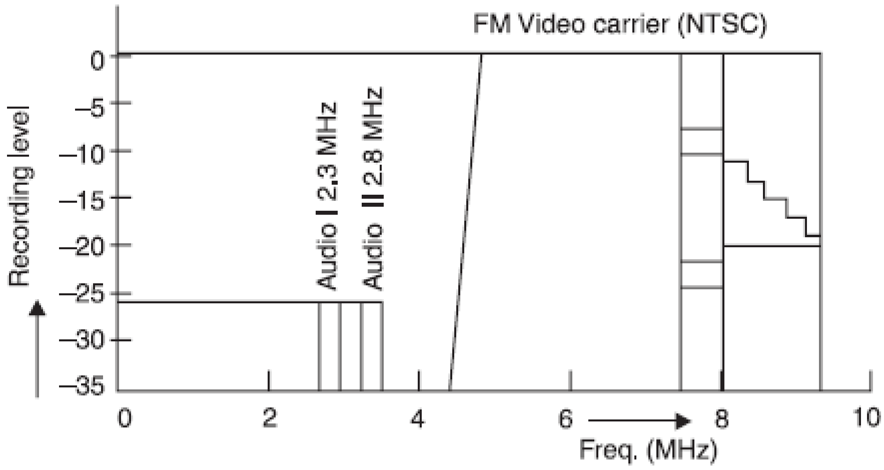


Fig. 34.14 The spectrum of video and audio signals. Black level is at 8.0 MHz white level at 9.2 MHz.

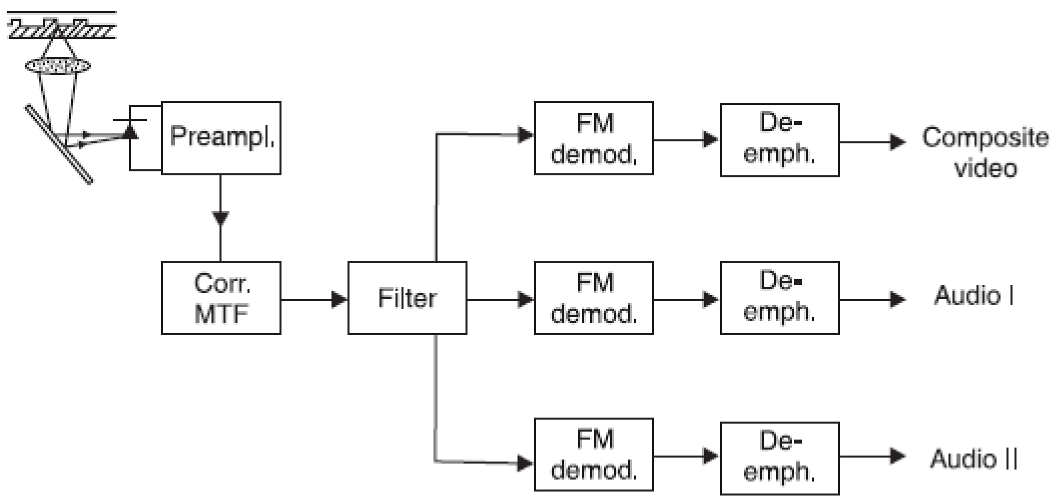


Fig. 34.15 Signal processing—decodings

DIGITAL ADDRESS SIGNALS

As part from video and audio information, the disc contains a number of signals, *inserted in the non-visible lines of the blanking periods.* These are in the first place *Vertical Interval Reference (VIR) and Vertical Interval Test (VIT) signals for testing purposes (lines 19,20,282,283) and secondly a* number of digital address signals for various purposes (lines 17, 18, 280, 281).

The *address signals* have the following functions:

A. *Lead-in tracks (min 600 tracks = 20 seconds)*
   Start code: The "read" objective is sent with normal speed forward to the program area on the disc.
   The "start" code is present in all fields of the lead-in tracks.
B. *Program area :*
   1. *Picture Code :* Consists of a picture number and a key for either continuous play in the normal forward mode or a stop key for automatic stopping in the "freeze frame" mode.
   2. *Chapter Code :* Consists of a chapter number and a key for either continuous search during depression of the search button or a stop key for automatic stopping of the search action. The chapter code is optional, depending on the type of program, and may be present only in the program area in those fields without a picture code.
C. *End tracks (min 600 tracks)*

   End Code : The read objective is sent back to the beginning with a speed 100 × normal speed.
   The end code is present in all fields of the end tracks.

SOLID-STATE LASER

When Philips originally invented its video disc system, the laser the company planned to use was an older, bulky type (Fig. 34.16) which was both *expensive and difficult to manufacture and align.*

Fig. 34.16 The original 'bulky' laser

As Philips went further down the track and invented the compact disc (CD) player, it became obvious that what was required was *a cheap, easily produced, robust laser system;* in other words, a solid-state laser.

The first commercially viable, and commercially available solid-state laser came from Matsushita, Japan's largest electronics concern, *MEL 4745 laser*. It was developed by the specialist semiconductor division of Matsushita, Matsushita Electronics Corporation. It is shown in Fig. 34.17.

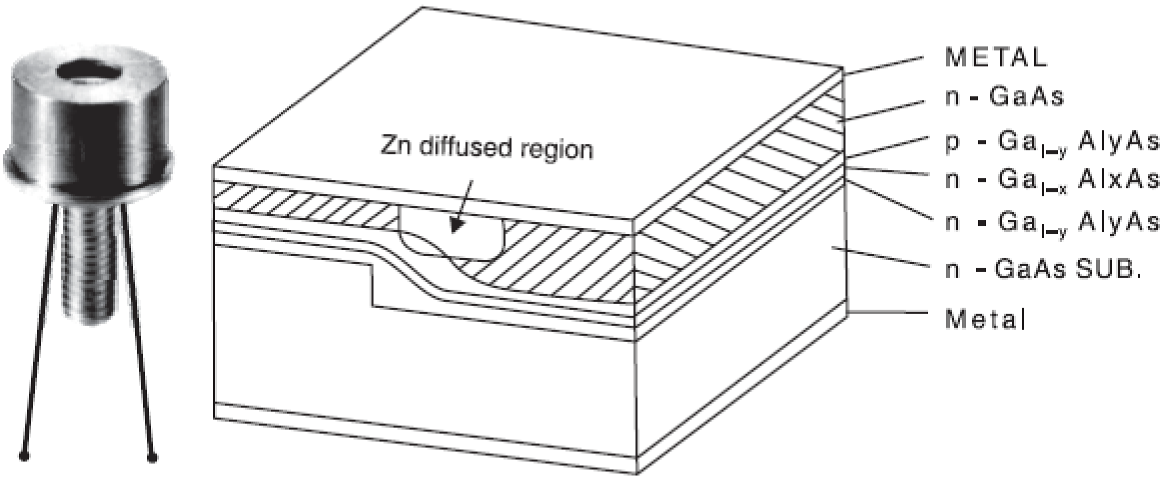


Fig. 34.17 Matsushita's terraced substrate (TS) semiconductor laser

In order for a semiconductor laser to be ideal for use in optical information processing systems, the following factors are essential:

1. Low threshold current
2. High power-output
3. Stable single transverse mode oscillation
4. Circular radiation pattern of laser beam
5. Long life

Conventional semiconductor lasers, however, could not satisfy all of these conditions. Matsushita succeeded in satisfying all of the above key factors by adopting a unique terraced substrate (TS) construction. *The TS laser can confine laser beams in a far narrower wave guide—resulting in higher power output with lower threshold current.*

In the fabrication of TS lasers, the face of the n-GaAs *substrate* is *terraced* by chemical etching, on which a double heterostructure is formed by successively growing an n-GaAlAs *cladding layer*, a GaAlAs *active layer*, a p-GaAlAs cladding layer and an n-GaAlAs *isolation layer.*

The active layer is grown thick at the terraced part, forming a modified rib—waveguide structure. *Such a structure is very effective for confinement of a laser beam and for a stable single transverse mode oscillation.* The TS structure also improves the concentration of injection current in the active region resulting in a remarkable decrease in *non-effective current.* The ratio of the *effective current* flowing through the terraced active region to the total injection current is in excess of 80 per cent.

Through these improvements, the threshold current has been reduced to a level of lower than 30 per cent of conventional lasers while still keeping a high power output.

The *TS laser MEL 4745* has a built-in PIN photodiode and can be easily connected with an APC circuit to keep the light output constant.

Characteristic Features–TS Laser MEL 4745	
Threshold current	30 mA
Light output	5 mW
Laser wavelength	810 nm
Radiation angle	13 degree (parallel)
	30 degree (perpendicular)
Mode	Signal transverse mode oscillation
Built-in PIN photodiode.	

FOCUSING

A thick video disc with a thickness of 1.1 mm rotating in free air at 30 Hz is generally *stable but not critically damped.* In order to restrict the action of transverse waves in the frequency range 0–15 Hz and also to compensate for the “umbrella” form deflection of the disc under its own weight, a *stablisation system* has been incorporated. *Stabilisation is achieved in the player by means of an air bearing, which is created by the centrifugal action of the rotating disc on the surrounding air.*

A plate with a central hole is positioned in the vicinity of the rotating disc's surface. This creates an air bearing having both the desired damping and stiffness characteristic. Disc stabilisation obtained in this way is by no means sufficient to guarantee the required accuracy on the *distance* between the disc and read objective.

In order to read the microscopically small information details on the disc, an objective lens with a large numerical aperture is required, which thus has a small depth of focus. *Numerical aperture (N.A) is defined as the product of the refraction index and the sine of the angle between the optical axis and the outermost light ray contributing to the imaging* (Fig. 34.18). With N.A = 0.4, a maximum *out-of-focus* allowed is 2 μm. In view of tolerances in the disc and player construction this accuracy can only be realised by means of a *servo control system* including a *moving read objective* that can follow the undulations of the disc.

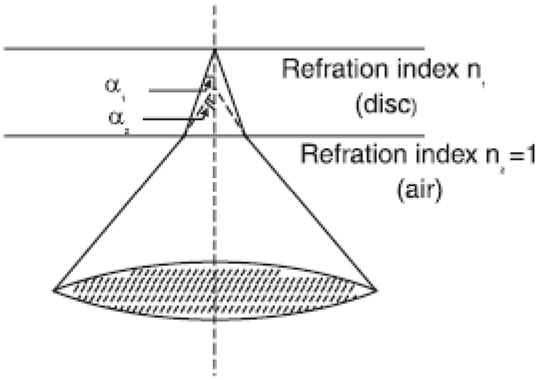


Fig. 34.18 Numerical aperture NA = $n_1 \sin \alpha_1 = n_2 \sin \alpha_2$

As can be seen in Fig. 34.19 the objective lens is mounted in a system similar to that of a loudspeaker voice coil and thus operates according to the *electrodynamic principle.* Depending on the direction and magnitude of current through the coil it executes *controlled vertical movements* in order to follow the irregularities.

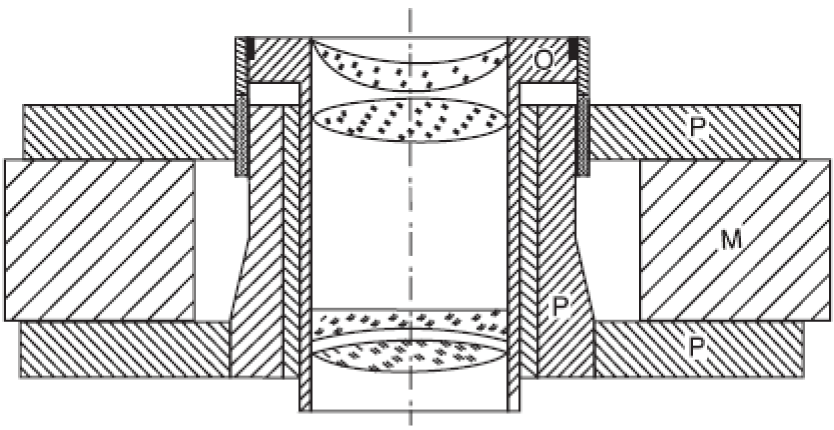


Fig. 34.19 Principle sketch of moving objective

To understand the way in which the *control signal* is obtained, it is necessary first to have a closer look at the photodiode which detects the light beam reflected by the disc. In fact this diode consists of two segments E and F and a third one between the two which is composed of four quadrants A, B C and D (Fig. 34.20). The light beam focused on the central segment will normally create a *circular spot* and thus all four segments will receive an equal amount of light.

The sum of the electrical signals over these four quadrants is the RF signal or video-information.

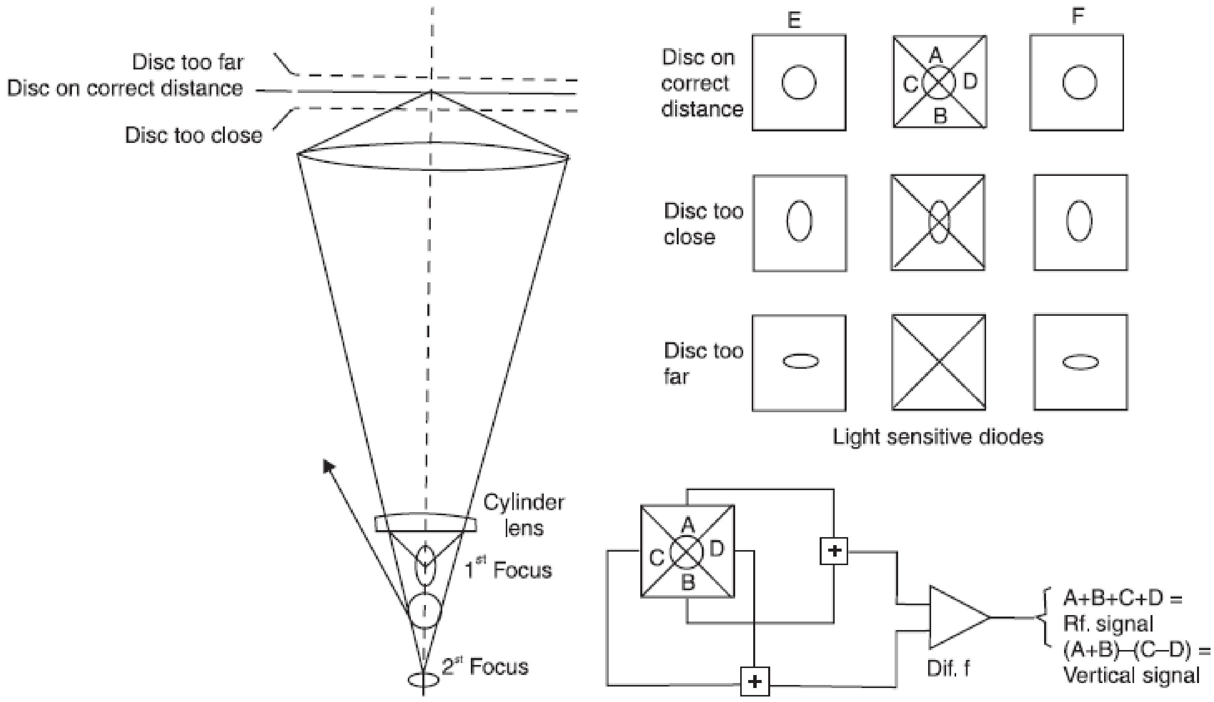


Fig. 34.20 Principle of astigmatic focusing

However, on its way to the diode the reflected beam passes a *cylindrical (astigmatic) lens.* As a result, the light spot on the diode which normally has a circular shape obtains an *elliptical deformation* if the information plane in the disc changes its distance from the objective. If this occurs each quadrant no longer receives an equal amount of light and so a difference signal can be derived. *This difference signal, once it has been amplified and processed, can be used to change the position of the objective lens with respect to the disc.*

The largest vertical movements of the disc occur at the speed of rotation and decrease rapidly at higher frequencies, Disc specification sets the maximum value at 10 g but the player can cope with accelerations as high as 14 g and still remain below an out-of-focus value of 2 µm.

RADIAL AND TANGENTIAL TRACKING

The information on the disc is contained in a *spiral that is read from the inside to the outside.* For this purpose the 'read' objective and other optical elements are mounted on a *sledge* driven by a small dc motor, and moving radially under the disc. With an average track pitch of 1.6 µm and the disc rotating at 1800 r.p.m. this means an average linear speed of the sledge of 3 mm/m.

The scanning light beam has to remain focused on the track with a *radial accuracy* of 0.1 µm, a requirement that cannot be met by a purely mechanical guidance system. By varying the speed of the drive motor by incorporating it in a *servo control system* certain slow corrections are possible. However, to cope with the effects caused by eccentricity of the spinning disc, for example, certain additional measures are required and use is made of *pivoting mirror* by means of which the light spot can move radially over the disc. This mirror is mounted in an assembly resembling a moving coil ammeter where the coil is part of the radial servo control circuit.

The *information* on the track can only be read optically and the *deviation* can also only be measured optically. For this purpose two *auxiliary beams* of light are used which are *slightly displaced* from the centre-line of the track, in opposite directions, so that they are *partly on* and *partly off* the track as shown in Fig. 34.21.

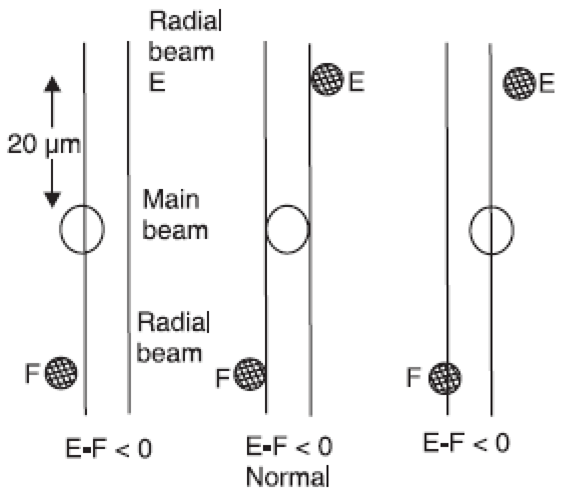


Fig. 34.21 Principle of track following

After deflection from the disc the two auxiliary beams each fall on its own photodiode E and F in Fig. 34.21 and *the average current through the diodes depends on the auxiliary beam relative to the track.* In fact the difference signal of the two diodes after amplification passes a low-pass filter with a cut off frequency of 20 kHz and is then used as an *error signal* in the control system. If the average position of the mirror deviates from its zero position, the average current is used to control the slide mirror for correction of the position of the sledge.

BLOCK DIAGRAM OF VLP PLAYER

From the discussion so far, the larger part of the block diagram will be self-explanatory. A few items, however, require an explanation.

In the event of a *dropout (an interruption of the signal due to any imperfection in the disc)* a drop-out detector operates as a switch by means of which *a parallel channel, fitted with a line scan delay, can provide information of the previous line.* Of the two available audio channels, either one of the two or both (in the case of stereo), can be connected to the VHF modulator. Since normal television receivers are monophonic, the two demodulated audio signals are available at output sockets at the back of the VLP player. These can be connected to a stereo amplifier.

BLOCK DIAGRAM OF VLP PLAYER

From the discussion so far, the larger part of the block diagram will be self-explanatory. A few items, however, require an explanation.

In the event of a *dropout (an interruption of the signal due to any imperfection in the disc)* a drop-out detector operates as a switch by means of which *a parallel channel, fitted with a line scan delay, can provide information of the previous line.* Of the two available audio channels, either one of the two or both (in the case of stereo), can be connected to the VHF modulator. Since normal television receivers are monophonic, the two demodulated audio signals are available at output sockets at the back of the VLP player. These can be connected to a stereo amplifier.

AUTOMATIC ACTIONS

A great many automatic functions have been incorporated in the VLP player both to ensure a *flawless performance* of the apparatus under all circumstances and in order to make it *foolproof.* A number of other automatic functions are particularly intended for making it *user friendly.* These are :

Priority mode	:	After switching ON the apparatus will always be in the normal play forward mode with both sound channels connected.
Program end	:	At the end of a disc the ‘read’ objective will automatically move with ‘search speed’ to the beginning of the disc. The same will happen after interruption of the playing of a disc by releasing the lid.
Program Start	:	If the ‘read’ objective lens arrives at the beginning of a disc and the lid is closed, the player will start again in the normal ‘play forward’ mode.
Stop during normal play	:	If an *appropriate picture code* is on the disc the player stops immediately at that picture and remains in the ‘freeze frame’ position.
Stop during search	:	If an *appropriate chapter code* is on the disc, the player stops the search action at the beginning of a chapter. Search is resumed after releasing and subsequent depressing of the ‘search’ button.

OPTICAL MEMORY DISCS

Magnetic discs are based on ideas which were current 30 years ago in the recording industry. With a normal record you start playing on the outside run and the track goes round and round in a spiral to the centre. One problem with the cassettes is that you have to record all the information in a *serial form.* So if you have recorded say 8 programs, and want the information stored in the last one, you have got to start all the way from the beginning until you get to the information you want. One way of overcoming this setback is to take a *normal LP-type record approach*, where although you have got serial recording format, you have designated points called *tracks.* As there is a certain amount of random access, it is possible to 'drop in' at any point.

A magnetic disc, Fig. 34.24, is divided up into *discrete tracks* in the same way as a normal LP record. But instead of the tracks being in a spiral *they are all concentric*. A small magnetic recording head, just like a recording head on a small cassette recorder, is then moved from say track 0 to, say, track 35 (Fig. 34.25). The disc unit is instructed that the computer wants to record the next chunk of information on track 5, and it moves the head to track 5 and dumps the information in a similar way to how the information was recorded on a cassette tape, i.e. by modulating the 1's and 0's pulses with a frequency that can be recorded magnetically.

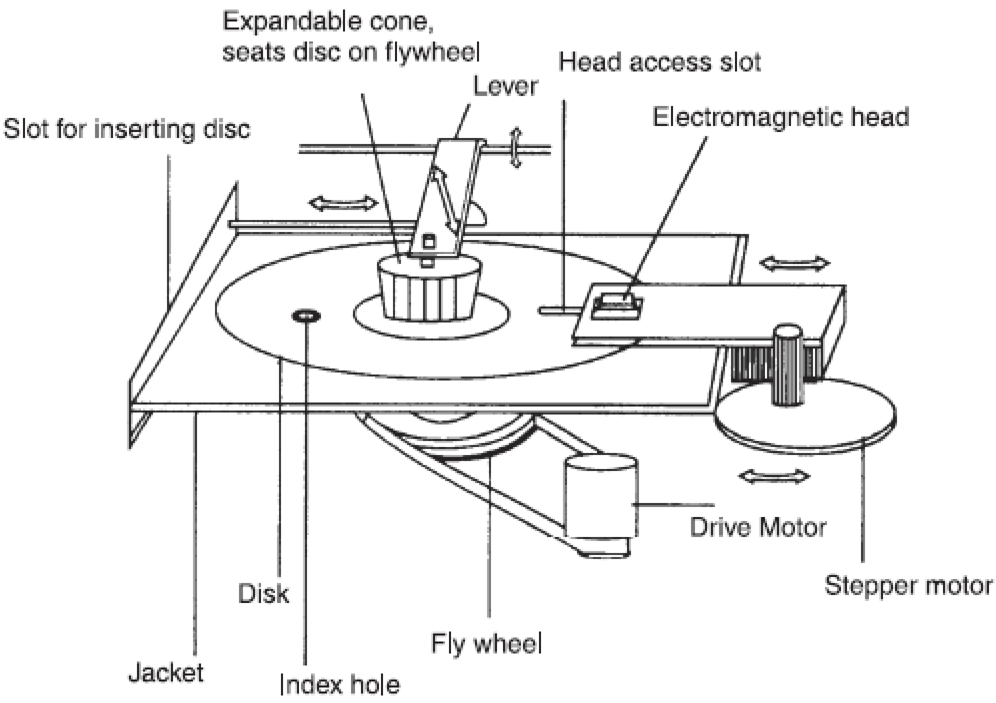


Fig. 34.24 Floppy disc and floppy disc drive

A few years ago IBM came along with the *floppy disc*, which is now being made in a whole series of difference formats, 8", 5.25" and now 3.5" in diameter. If a sheet of plastic covered with a thin layer of magnetic material is spun at 300 r.p.m. it keeps relatively

The disc is split into a *series of tracks*, which vary from about 35 to 80. Each track is itself split into a number of *sectors.*

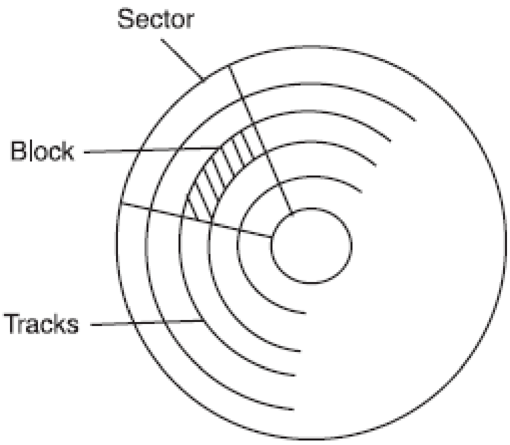


Fig. 34.25 How data is recorded on a floppy disc

The number of sectors vary, but it is usually about 16. In other words the disc is made of a *series of blocks.* So to store information it must be possible to keep track of which block or group of blocks the information is stored. One of these tracks is designated the *index track,* and on that track you store an index of what information is on the disc and where it is. It's like looking up something in a book; there is an index which tells the recording head exactly *where* all the information is stored and usually *how much* information is there. There are many more sophisticated methods where you might have various check digits or other information stored on the index tracks.

Anyone who has used *disc drives* will have heard it *whirring away* making clunking noises. This is the head trying to find the appropriate track. The *size* of each block can vary from between 256 bytes to 1024 bytes. *Unlikes a cassette recorder the information can be spread all over the place.*

EXERCISES

Descriptive Questions

1. Compare a video disc with a gramophone record.
2. What are the similarities between the three video disc formats?
3. Briefly explain the different optical recording mediums.
4. With the help of a suitable sketch explain the construction of a video disc along with the dimensions of
    a. diameter
    b. track pitch
    c. pit width
    d. pit depth
    e. thickness
    f. protection layer
5. Briefly explain the recording and playback system of an optical video disc.
6. Describe the digital address signals.
7. Explain the working of a TS laser.

8. What is the significance of radial tracking?
9. What are the different controls on a VLP player?
10. What is the difference between a video disc and an optical memory disc?
11. How data is recorded on a floppy disc?
12. Explain the following :
    a. serial form
    b. discrete tracks
    c. blocks
    d. index tracks

Multiple Choice Questions

1. The three video disc formats are
    a. interchangeable
    b. non-interchangeable
2. With the video disc as information carrier one
    a. does not have immediate access to any program part
    b. has immediate access to any program part
3. The information on the video disc exists as a
    a. spiral track starting from outside
    b. groove starting from inside
    c. spiral track starting from inside
    d. groove starting from outside
4. There are two types of laservision discs
    a. standard play
    b. long play
    c. esxtended play
    d. active play
5. The RCA stylus recovers information from the disc surface by
    a. indirect electrical means
    b. direct electrical means
6. The TeD and RCA systems
    a. function even if the disc surface is damaged
    b. fail to function if the disc surface is damaged
7. In response to a tracking error signal the stylus in a VHD system can move
    a. longitudinally
    b. transversely
    c. longitudinally and transversely
8. With a single information track, the information is encoded in
    a. the length of the pits
    b. the spacing of the pits
    c. the length and spacing of the pits
9. The old bulky lasers were
    a. expensive
    b. difficult to manufacture
    c. difficult to align
    d. all of the above

10. The TS laser has
    a. higher power output with lower threshold current
    b. lower power output with higher threshold current

Fill in the Blanks

1. The superiority of the Laser Disc Player originates from the application of _______ for reading out the information on the disc.
2. There is an effective _______ of information on the disc against finger prints, dust and scratches.
3. Average track pitch is _______.
4. Pit width is _______ whereas pit depth is _______.
5. In practice, however, _______ magnetic recording tape is used as a _______ carrier.
6. The laser optical system is also called _______.
7. There are two variations of the laser optical system, _______ and _______.
8. Long play discs are for straight forward _______ playback of entertainment program.
9. Active play discs involve the viewer in learning a new _______.
10. _______ is the cheapest and the most basic of the three video disc formats.

ANSWERS

Multiple Choice Questions

1.	(b)
2.	(b)
3.	(c)
4.	(b)
5.	(a)
6.	(b)
7.	(c)
8.	(c)
9.	(a) & (c)
10.	(a)

Fill in the Blank

1.	optics
2.	protection
3.	1.6 μm
4.	0.4 μm, 0.1 μm
5.	50 mm, program
6.	Philips VLP system
7.	reflective, transmissive
8.	uninterrupted
9.	skill
10.	Selectavision(RCA)

CHAPTER 35

REMOTE CONTROLS

One convenient feature of modern TV receivers is their remote control operation. Remote control, as the name implies, is a system by which one is able to control the performance of an equipment placed at a distance.

Many television receivers have provisions for operating the ON-OFF, volume, main tuning, picture contrast and brightness adjustments remotely from a normal viewing distance. The command signal travels to the set in a variety of ways depending on the system. By far, the most popular remote control systems are those which employ ultrasonic signals. The frequencies commonly used for remote control lie in the 40 kHz range.

The remote system has three main components—a remote transmitter, a remote receiver and the devices at the receiver to effect the change ordered by the remote command. These include motors, relays and, in the latest systems, electronic control devices.

ULTRASONIC TRANSDUCERS

The operation of an acoustic type remote control system depends upon the use of *transducers*. One transducer in the transmitter, i.e. in the hand-held remote control unit, converts a mechanical vibration into high-frequency sounds. Another transducer in the receiver converts the high-frequency sounds into electrical signals which can be used for operating a control in the receiver.

Two basic transmitter types are used. The first is the mechanical or "bonger type" which employs a cylindrical aluminium rod which vibrates in its longitudinal mode when struck. If such a rod is struck on one end by a hammer moving along its axis, it emits a sustained note which has a definite frequency. For example, an aluminium rod 2½" long has a fundamental resonant frequency of about 40 kHz. The internal damping of the aluminium rod is so slight that a large part of the vibrational energies stored in the rod, after the original blow of the hammer, is radiated.

A single rod of a specified length will produce a certain resonant frequency. Thus, to control three or four functions within a television receiver, three or four rods of slightly different lengths have to be used. In a typical transmitter, the working parts include a hammer and a steel cylinder weighing about 2½ grams located at one end of the cylindrical rod. When a button is pressed, the hammer is pushed away from the rod by the force of a spring. As the button is pressed further, the spring is suddenly released and the hammer strikes the rod. Usually, if more than one hammer is employed, the difference in frequencies of the various rods may be to the tune of 1000 Hz or more.

Once the rod is struck and the energy transmitted, it is essential to dampen the remaining energy as quickly as possible. For this purpose, a *mechanical damping* method is employed. In this method, damping is achieved by a small piece of spring wire, covered by a plastic sleeve, which protrudes through the mounting and touches the rod. When the button is pressed, the damper is withdrawn and when it is released, the sleeve again makes contact with the rod.

In the receiver, the ultrasonic energy is picked up by a crystal type microphone using a barium titanate crystal element. This material, when cut in the form of a small bar or plate, is mechanically strained by the sound waves. Conversely, when a voltage is applied across the two electrodes, the barium titanate gets mechanically strained. This is the well known *piezoelectric effect.*

In Fig. 35.1 (a) two thin rectangular wafers of barium titanate are combined together. The conducting electrode silver, is applied at both the ends of the wafers. These electrodes are shown by the dark portion on both ends of the assembly. Between the two wafers, at the nodal points of vibration, two thin metal strips are cemented which serve as electrical contacts and as a mechanical suspension.

To broaden the response of this transducer, a small U-shaped piece of aluminium is added to the assembly, Fig. 35.1 (b). This makes the microphone behave like *two tuned circuits coupled together*. It enables ultrasonic frequencies from 37.75 kHz to 41.25 kHz to be picked up and converted into their equivalent electrical signals.

Figure 35.1 (c) shows a cross-section of the entire microphone. On the left end, a supporting piece is present which carries the barium titanate wafers on the two thin metal strips mentioned earlier. Next comes the aluminium bridge, to which the wafers are joined together with a plastic piece, and a rectangular window which fits closely around it. Beyond the bridge, there exists a space equal in length to a quarter of wavelength at about 40 kHz. Finally, after this space, comes a two inches long rectangular horn.

A *quarter-wave section* may be thought of as a transformer to match a load of Z_R ohms to a source of Z_S ohms. Such a match can be obtained if the characteristic impedance *R*o of the quarter wave section satisfies the relation $Ro = \sqrt{Z_S\ Z_R}$.

Both the one-quarter wavelength space and the horn serve to match the impedance of the barium titanate assembly to the air. In other words, the horn and the one-quarter wavelength space help to couple the wafers to the air.

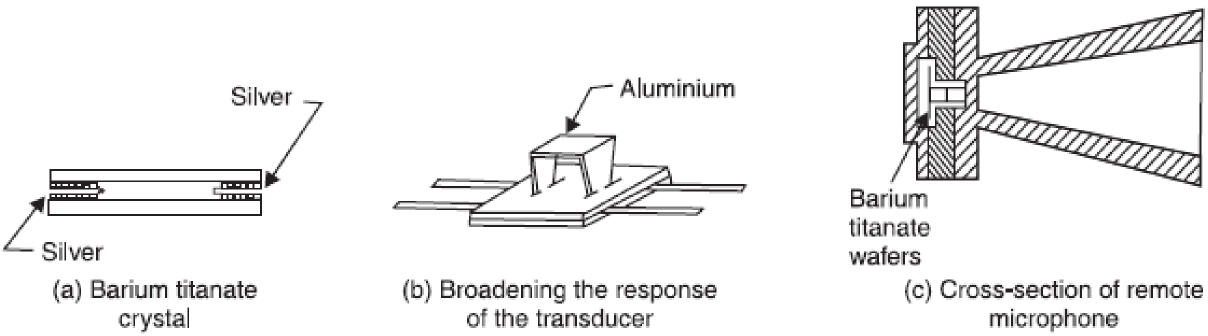


Fig. 35.1 Microphone used to pick-up ultrasonic sound of transmitter

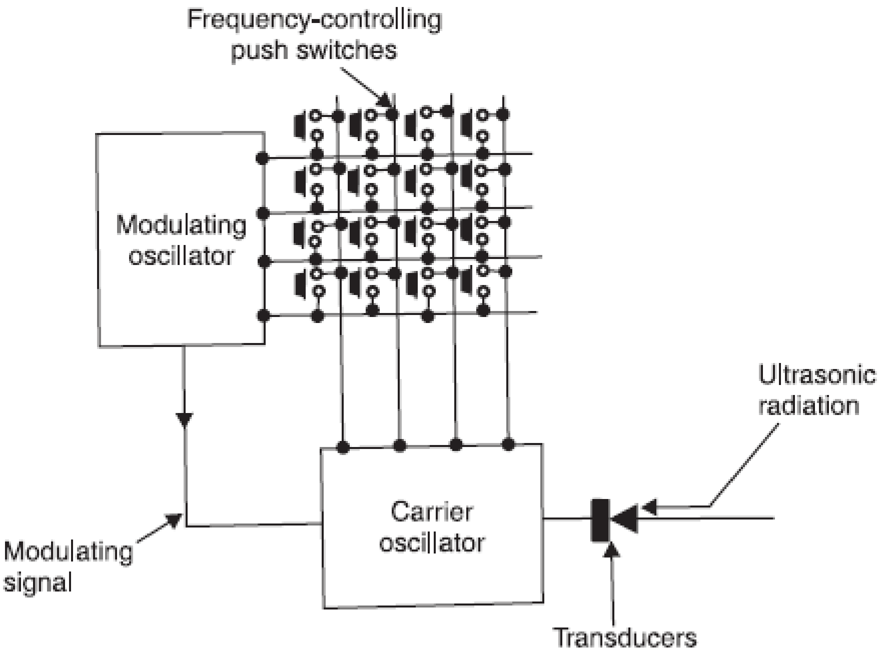
The combination of the mechanical transmitter, the microphone and the amplifier in this system, provides enough sensitivity to make aiming of the transmitter quite unnecessary in most of the horns. The sound reflected from the floor, walls, ceiling or furniture makes it possible to operate the receiver controls with the transmitter in almost any desired position —

held in the hand or resting on a table or chair. The *line of sight and the approximate aiming of the transmitter* become important only at the maximum range—a distance of 40 feet — which is rarely needed.

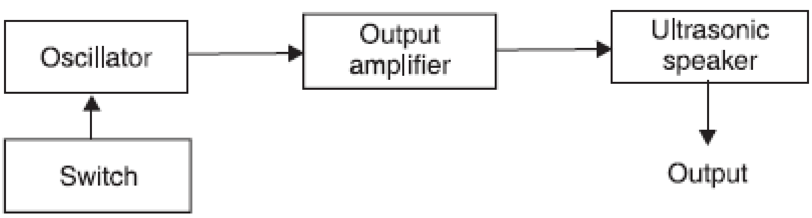
FREQUENCY SIGNAL ENCODING

The receiving part of the remote control system receives the particular command signal transmitted by the remote transmitter and the required operation is carried out. The receiver contains a crystal matched to that in the transmitter and also a number of tuned circuits which respond to the different frequencies transmitted. The ultrasonic signal is sensitive to external noise. To overcome this drawback, different methods of encoding the signal can be used. One of them is the *frequency signal encoding.*

In the system, Fig. 35.2 (a) the transmitter uses two separate oscillator. One of the oscillator generates a carrier wave and the other, a modulating signal. By a slight variation in the tuning circuit, the carrier oscillator can produce four or more different frequencies, e.g. 35.43 kHz, 38.29 kHz, 41.15 kHz and 44.01 kHz. Similarly, the other oscillator, for generating modulating signal, can produce 148 Hz, 193 Hz, 254 Hz or 333 Hz. A modulating signal amplitude modulates any one of the four carrier frequencies depending upon the remote operation required. In this way, one would be able to select any of the 16 frequency combinations for transmission by a 4 × 4 matrix of push button switches. *Each frequency combination represents a specific instruction.*



(a) Frequency signal encoding



(b) Block diagram of electronic remote transmitter

Fig. 35.2 Ultrasonic transmitter

In the ultrasonic receiver (Fig. 35.3), a *narrow bandpass filter* (F1) allows only those signals whose frquencies lie within 35 kHz to 45 kHz range. All such signals are then applied to four more narrow bandpass filters (F2 to F5), only one of which will allow the frequency, already accepted by F1, to pass. Thus, these filters act as signal sorts ensuring that the desired signal is always routed to its correct terminal on the *matrix decoder.*

On leaving F2–F5, the desired signal is rectified by one of the four detectors D1–D4. It is then applied without H.F. filtering to any one of the *detector-integrators* D5 to D8, and to the signal-combiner unit. Since a detector integrator can produce an output only if it receives the correct input, a rectified signal at the input of a particular detector-integrator produces a dc signal at its output. This signal is applied to the appropriate terminal of the matrix decoder as a *carrier identifier*. It is now necessary to determine which of the four possible modulating frequencies is being applied to the identified carrier. This is the purpose of the *signal combiner* and of the four filters F6 to F9.

The signal combiner accepts any of the rectified, but unfiltered, modulated carriers derived from detectors D1–D4 and feeds the one it receives simultaneously to the four narrow-bandpass filters F6–F9. Only one of these filters will allow the modulating signal of the received carrier to pass, and so only one of the four possible modulating signals will be allowed to reach the detector integrator, D9–D12.

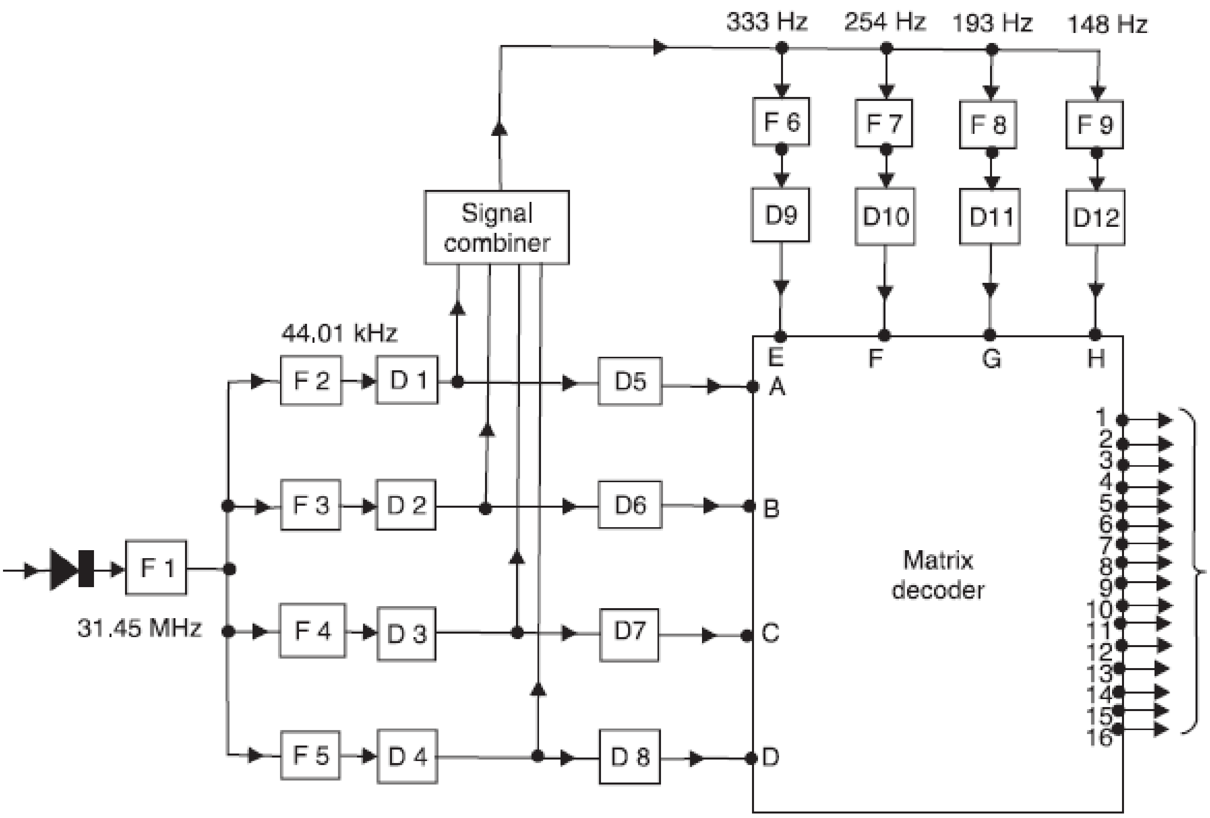


Fig. 35.3 Ultrasonic receiver

These operate in the same way as D5–D8, producing at their outputs dc signals corresponding to one particular modulating frequency. These dc signals are applied to the appropriate terminals of the matrix decoder as *modulation identifiers.* From these two identifiers the matrix decoder produces a signal at one of its 16 output terminals.

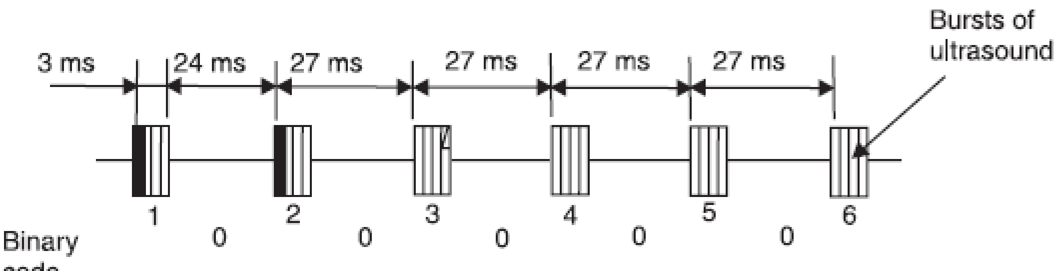
PULSE POSITION MODULATION ENCODING

Pulse position modulation (PPM) encoding is the system where encoding is carried out by positioning the pulses. In the PPM method of signal encoding, *the transmitted signal takes the form of a binary coded word formed from a succession of short bursts of ultrasound.* Ultrasound frequencies are not included in the code. The code is determined by the time interval or space between successive ultrasound bursts. This principle is illustrated in Fig. 35.4. Here the frequency of ultrasound is 40 kHz.

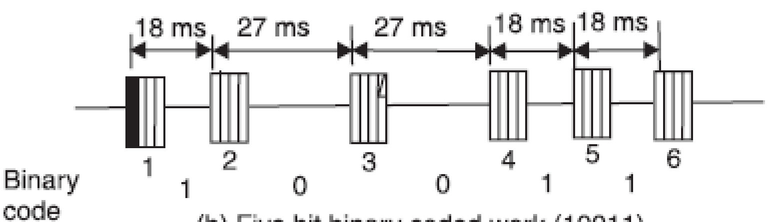
In Fig. 35.4 (a), a regular train of six ultrasonic pulses, each of duration 3 ms, is repeated at 27 ms intervals. The *separation rate* is therfore 1000/27 or 37 Hz and the separaton time from trailing edge to leading edge is 24 ms. Suppose that the pulse separation interval of less than 20 ms is recognised by the receiver as binary-1 and more than 20 ms recognised as binary-0. Six equally spaced ultrasonic bursts with spacing interval greater than 20 ms represent five-bit binary code 0 and therefore identified by the notation 00000.

In Fig. 35.4 (b), five-bit binary coded word is identified by notation 10011 since the first, fourth and fifth pulse spacings are *less than* 20 ms and the second and third spacings are *more than* 20 ms.

For example, the PLESSY IC SL490 is a low power remote-control transmitter capable of driving a fixed frequency ultrasonic transducer by using the technique of PPM for signal encoding, and an 8 × 4 push switch matrix to provide up to 32 commands.



(a) Regular train of ultrasonic pulses



(b) Five bit binary coded work (10011)

Fig. 35.4 Pulse position modulation (PPM) encoding

ENCODING BY TIME RATIO DISCRIMINATION

Time ratio discrimination (TRD) is also known as *pulse width modulation* (PWM). Despite coding taken as spacing between the pulses, *the width of the pulses is taken here for identification of binary notation.* Therefore, every bit of transmitted word is allocated a period of time within which the burst of ultrasound representing the bit must be transmitted. The actual duration of the transmitted bit is then expressed as a percentage of the allocated bit period, and the different percentages which result are used to represent binary-0 and binary-1.

The Philips SAA–05000A type remote-control transmitter is based on *time ratio discrimination* type of signal encoding. The principle of TRD is illustrated in Fig. 35.5. For example, binary-1 is represented by a burst of ultrasound whose duration is 66.6% of the total bit period. The binary-0 is represented by a burst duration of 16.6% of the total bit period.

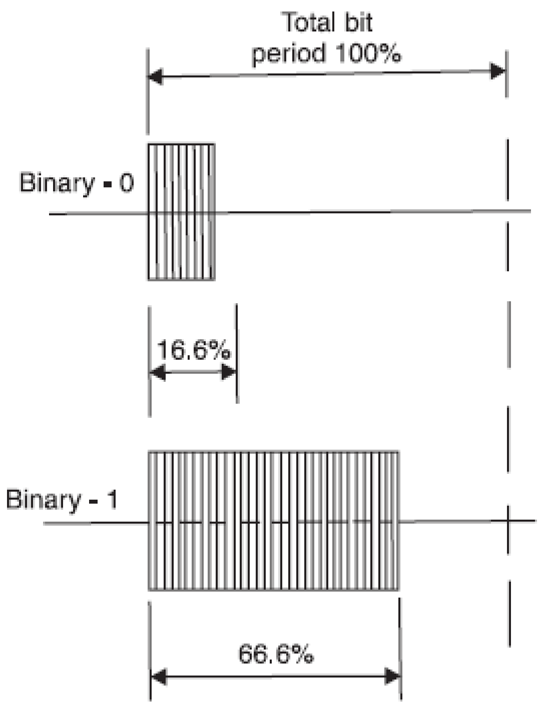


Fig. 35.5 Encoding by time ratio discrimination (TRD)

REMOTE CONTROL TRANSMITTER

The SAA–5000A is a MOS N-channel IC which provides the *encoding and modulation* function for the remote control of television receivers, including those equipped with Teletext and Viewdata facilities. The internal block diagram of IC–SAA 5000A is shown in Fig. 35.6. The transmitter using SAA–5000A is based on the time ratio discrimination encoding system. It provides 32 commands which are activated by switch controls on the keyboard matrix. The IC is connected to this keyboard by six matrix outputs and six matrix inputs.

It utilises the TRD method of word encoding. It trasmits a 24 bit code for every transmitted command. It consists of a 7 bit "Start-Code" followed by a 5 bit "Message". At the end of the message code, all the 12 bits of "*start" and "message" code* are retransmitted but this time with their individual bits *inverted.* Thus, the total transmitted word consists of 24 sequential bits, the last 12 of which are mirror images or complements of the first 12. The entire transmitted 24 bit word is repeated till the switch of the keypads remains depressed.

The advantage of sending 12 bit "start" and "message" code followed by 12 bit *"complementary code"* is that the receiver is easily able to distinguish the required code from external noise. Figure 35.7 explains the TRD word-coding sequence. The message code shown is the binary number 10001, but it could be any one of 32 combinations in all. The IC SAA–5000A is suitable for *ultrasonic and infrared transmission modes.*

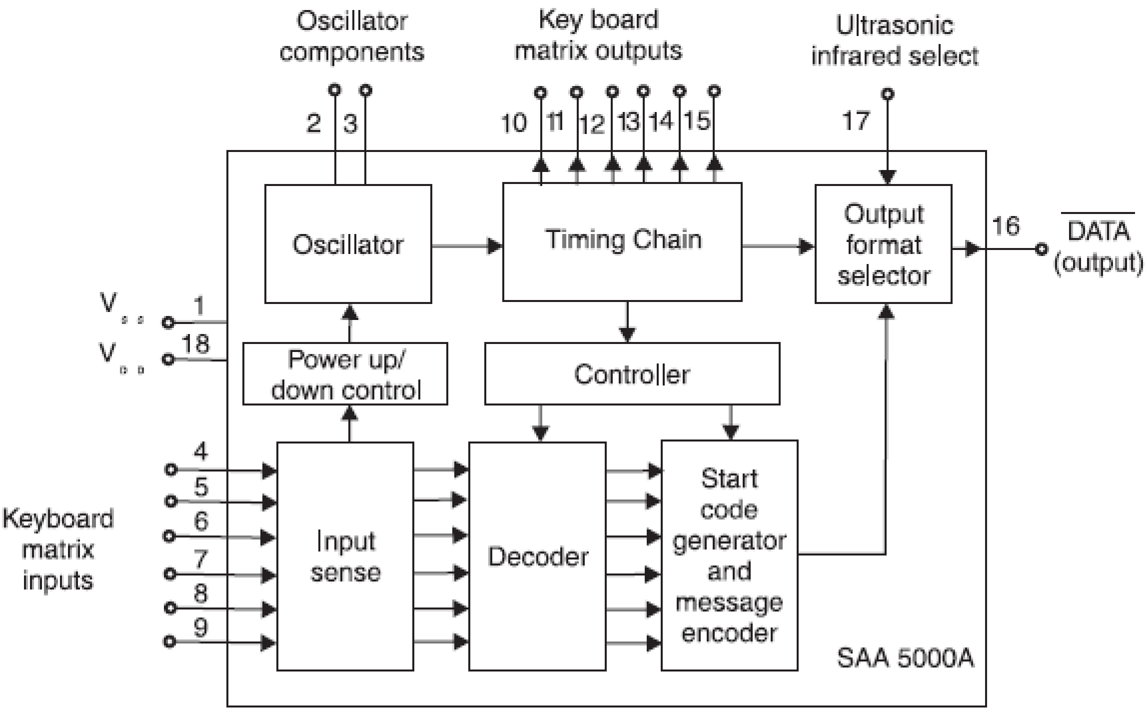


Fig. 35.6 Internal block diagram of IC SAA–5000A

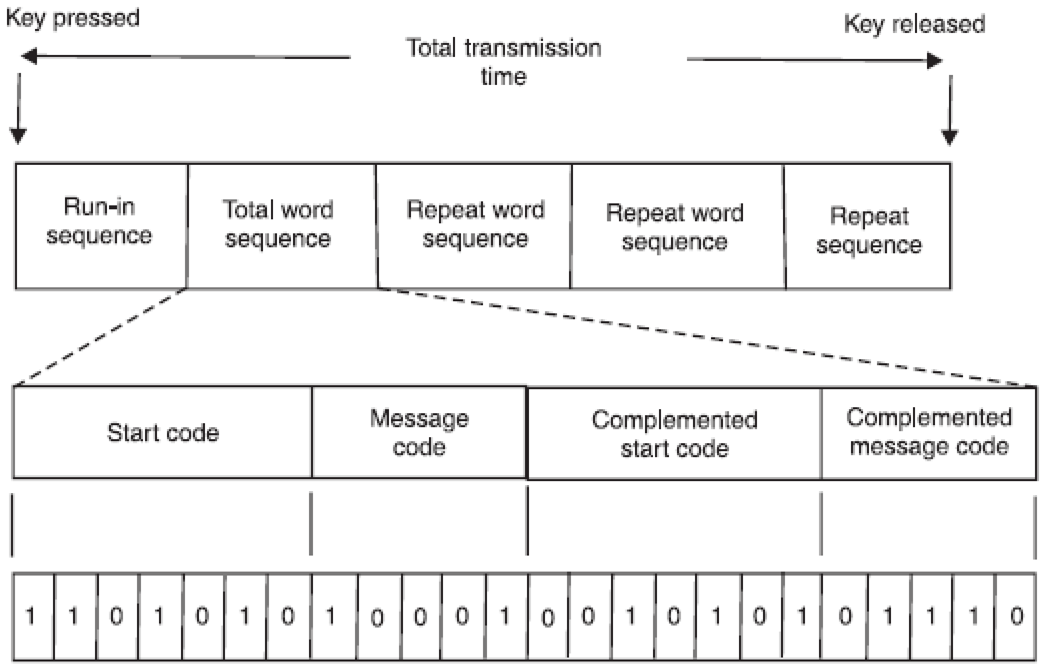


Fig. 35.7 24 bit command

DESCRIPTION OF ULTRASONIC TRANSMITTER CIRCUIT

The ultrasonic transmitter circuit is shown in Fig. 35.8. The transmitter accepts data from the 32-keypads and coverts it into a 24–bit TRD-encoded word. The encoded data appears as a sequential switching waveform at pin 16 of the IC. It is used to turn ON and OFF the stable multivibrator, formed by the two transistors Q1 and Q2 and the associated components.

When the multivibrator is turned ON, it operates at a frequency of 40 kHz and is used to drive the ultrasonic transducer.

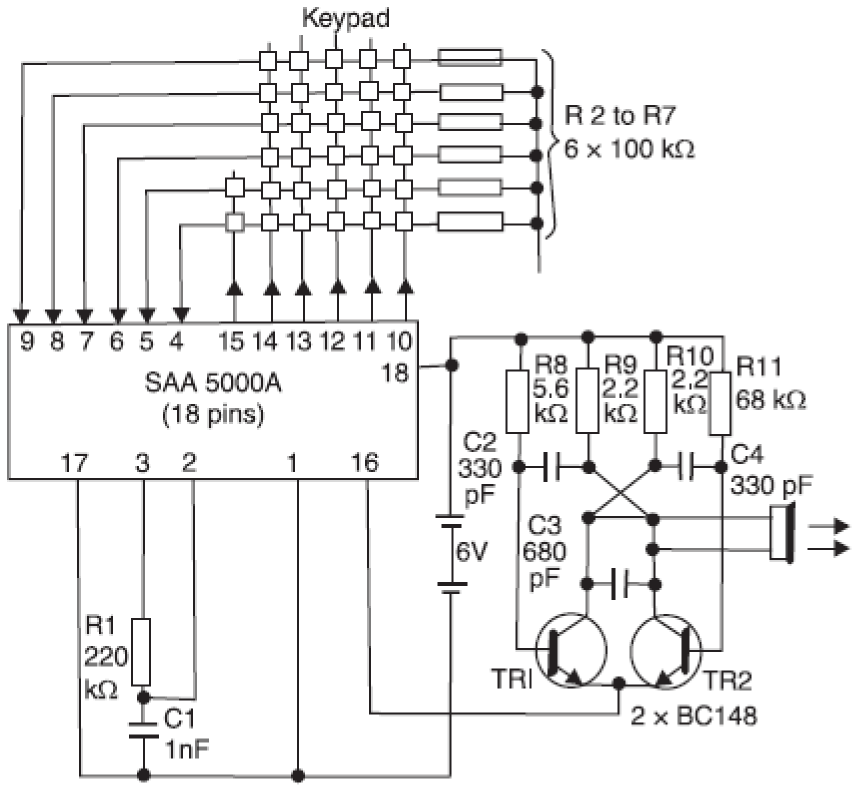


Fig. 35.8 Ultrasonic transmitter circuit

The bit period of the encoded word is determined by an oscillator within the IC whose operating frequency is set by the 220 kΩ resistor and 1000 pF capacitor connected externally to pins 2 and 3.

With the values shown in Fig. 35.8, the bit period would be about 82 ms but it can be varied over the range 0.8 ms to 10 ms. Pin 17 of the IC is shown connected to pin 1, which in turn, is connected to the negative electrode of the internal battery. This puts the IC into ultrasonic operating mode. A typical remote control system is illustrated in Fig. 35.9.

TROUBLE SHOOTING REMOTE CONTROL SYSTEMS

The trouble shooting assumption that *a defect is the cause of a symptom* is also used in servicing a remote system. The steps for isolating the defective component depend on the specific system. In general, the first step is to check whether the receiver operates in the *manual mode.* If it does, the trouble is in the remote control system. If the manual system is also non-operative, the fault lies with the set, in which case, the set should be serviced.

The remote system consists of a number of *sections* and/or *modules.* The transmitter can be checked by replacing it with another one in a working condition. If the transmitter is the cause of the trouble, check the battery and verify the frequencies. *Mechanical transmitters* are often damaged by rough handling and ideally should be replaced rather than repaired. If battery replacement does not help, check the components of the electronic transmitter. Since there are only one or two stages, part checking and replacement should help locate the defective part quickly.

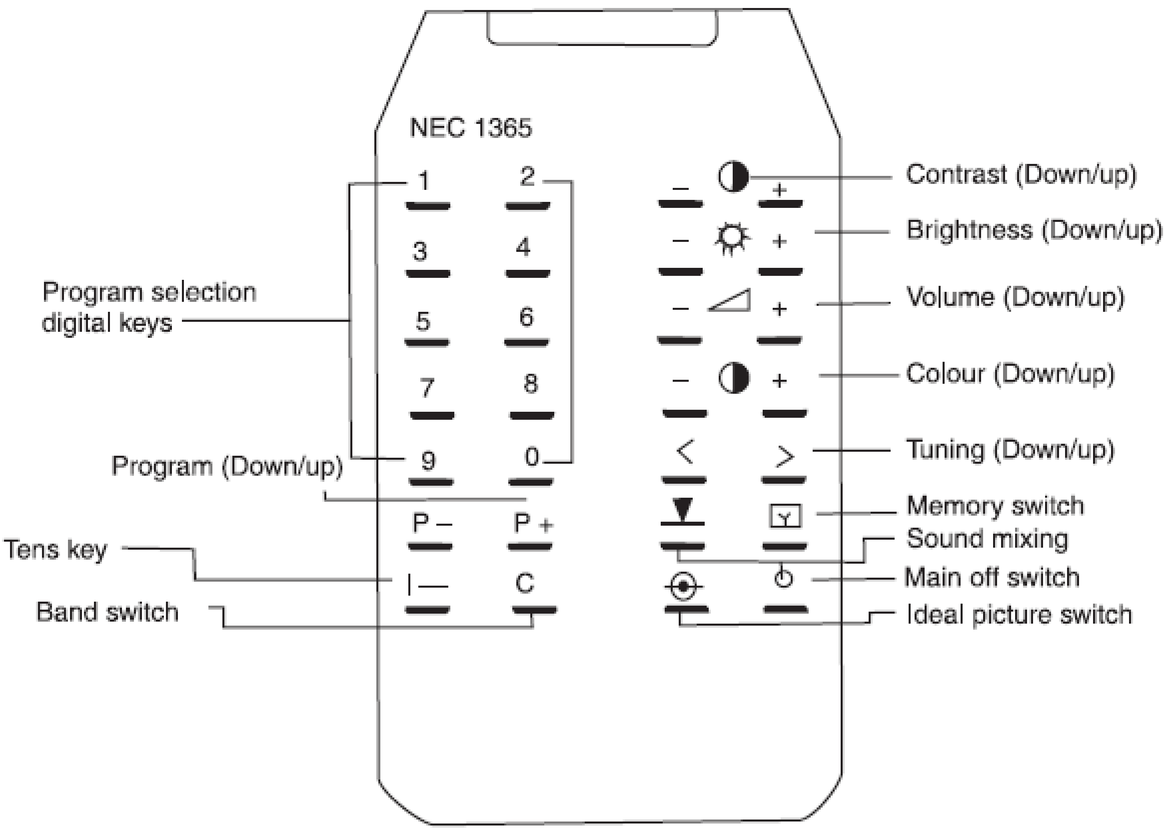


Fig. 35.9 Remote control system

If a receiver operates by manual control but not by the remote, and if even some other working transmitter does not operate it, then the trouble has to be in the remote receiver. In trouble-shooting, it is important to obtain as much information from the symptoms as possible. Thus, *all the remote functions should be tried.* If some functions operate and others do not, then the trouble probably lies in the connections or in the activating device for the inoperative functions. If none of the functions operate, this means the trouble is in a common section such as the transducer pick-up or the broad band amplifier. An accurate broad band audio generator can be used as a *signal substitute* to feed a signal directly to the remote receiver for signal-tracking. Here, a problem often overlooked is the possibility of overloading the remote receiver with some very strong, unrelated signal (e.g. the horizontal sweep signal). *The remote receiver must be located at the correct point in the chassis and all shielding must be in place.* If it seems that the horizontal signals are interfering with the remote control, *disable* the horizontal section and monitor the operation of the remote receiver by checking the voltage change at the output of the function that is being activated.

The technician must be familiar with the normal operation of the system. Some receivers have *preset positions* where colour, tint, brightness and contrast are set to a fixed value when the *automatic system* is activated. Under this condition, the remote tint control, for example, will have no effect. This is normal and not a defect. The automatic setting must be shut off in order to permit the remote functions to operate. As another example, in some digital systems the channel number must be entered from a keyboard, one digit at a time. Suppose if channel

7 is desired, then the digits 07 must be entered. If only the key 7 is pressed the receiver will automatically shut off after about 15 seconds. This *safety feature* prevents the receiver from being turned on by some accidental ultrasonic noise. Just as it is important to know the details of the system, so also it is necessary to teach the user *proper operating procedures* to prevent nuisance service calls caused by a lack of knowledge or by misunderstanding.

Most remote systems are modularised, allowing *module substitution* for quickly locating a defective module. A defective module can then be serviced in the shop or sent back to the factory for exchange. The use of modules requires many interconnections. Plugs and sockets do not always make a good contact. One simple remedy is to plug the module in and out several times. The *contact wiping action* resulting from this procedure may eliminate a poor connection. Cables are used for many interconnections. The *dress* of these cables may be critical to prevent feedback or pick-up. Moreover, these cables should not be permitted to rest on hot components because the insulation may melt. Repeatedly moving these cables can also cause individual wires to break, usually near the plug or socket. This type of problem may frequently be spotted by *visual inspection*, which must always be a part of the early service procedure.

The *proper position* of the microphone is important for a normal pick-up of ultrasonic energy. The microphone often fits into a housing on the front panel. It may be jarred loose by transportion or by carelessness after an unrelated repair. It is necessary to check all receiver functions after any repair.

The operation of a typical colour TV receiver remote control is illustrated in Fig. 35.10 *Note that the Mains Switch must be in the ON postion for the remote control unit to function.*

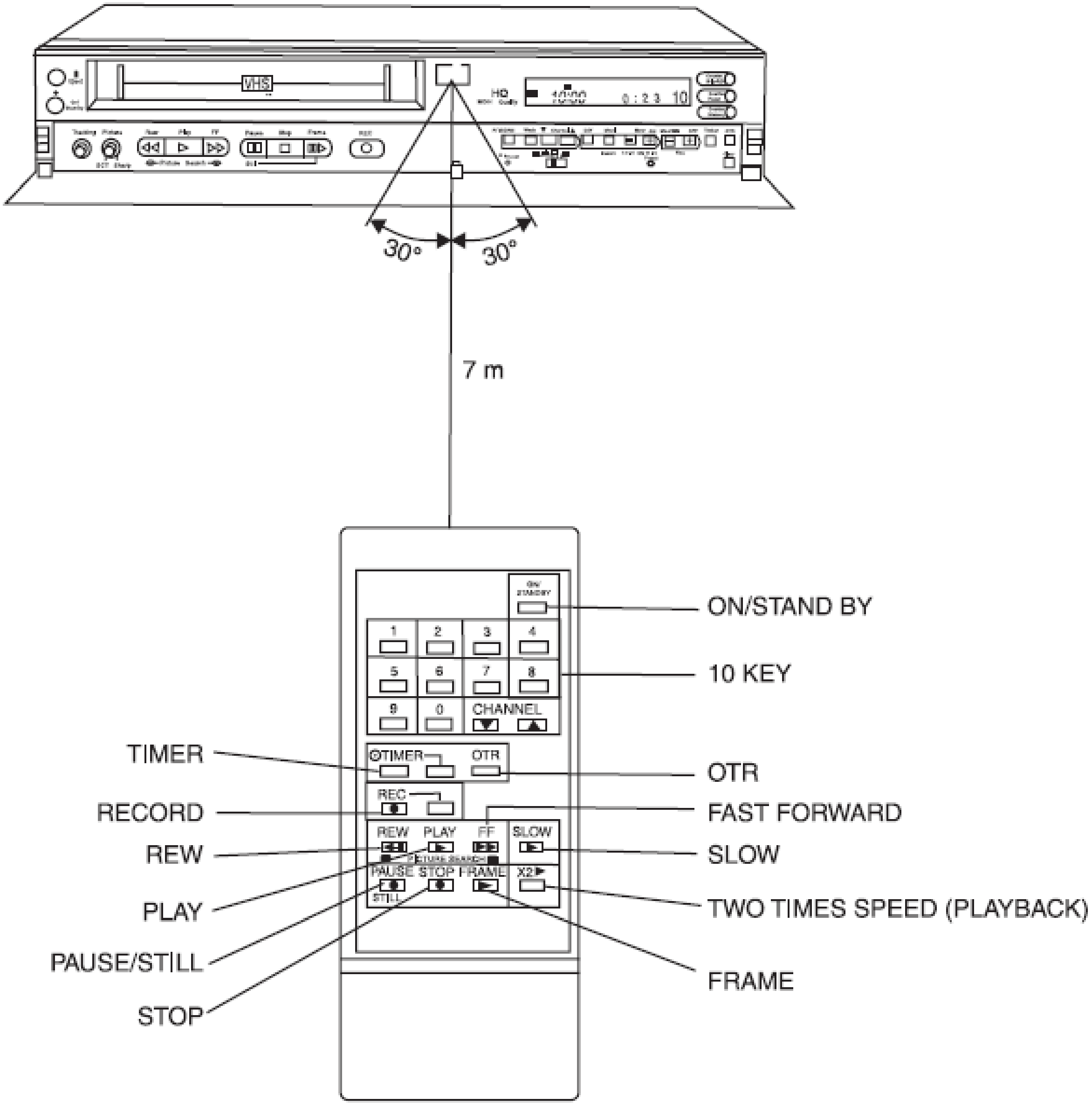


Fig. 35.10 Operation of a colour TV receiver remote control

REMOTE CONTROL OPERATION

Remote control button functions:

Each function of the remote control except for the SLOW, 10 key and two times speed playback function corresponds to those of the buttons provided on the front panel of the VTR. You can operate 27 functions in all by using the remote control unit which contain CHANNEL UP/DOWN, 10 key (direct channel selection), ON/STANDBY, RECORD, REWIND, FAST FORWARD, PICTURE SEARCH (FORWARD, REVERSE), PLAY, PAUSE/STILL, SLOW, STOP, FRAME, OTR, TIMER and two times speed playback modes.

The remote control must have “Line of Sight” to the INFRARED REMOTE control receiver on the front of the VTR. It must be within an angle of 30° either side of centre. When you use the remote control unit while aiming it at the front side of the VTR, the *maximum* operating distance is about 7 m (within 30° either side of centre, the *effective* operating distance is approximately 5 m).

EXERCISES

Descriptive Questions

1. Describe the working of an ultrasonic transducer.
2. Explain the working of a remote transmitter with the help of a block diagram.
3. Explain the working of an ultrasonic receiver.
4. Differentiate between PPM and TRD methods of encoding.

Fill in the Blanks

1. One convenient feature of TV receivers is their _______ operation.
2. Remote control is a system by which one is able to _______ the performance of an equipment placed at a _______.
3. Many TV receivers have provisions for operating the _______ remotely from a normal viewing distance.
4. The remote system has three main components a sremote transmitter, a remote receiver and the devices at the receiver to effect the change ordered by the remote _______.
5. The line of sight and the approximate _______ of the transmitter become important only at the maximum range.
6. Each frequency combination represents a specific _______.
7. In the PPM method of signal encoding the transmitted signal takes the form of a _______ coded word formed from a succession of short bursts of _______ .
8. In TRD the _______ of the pulses is taken for identification of binary notation.

ANSWERS

Fill in the Blanks

1.	remote
2.	control, distance
3.	ON-OFF, volume, maintuning, picture contrast, and brightness adjustments
4.	command
5.	aiming
6.	instruction
7.	binary, ultrasound
8.	width

CHAPTER 36

VIDEO SYSTEMS

Video systems have become an essential part of the household entertainment. There are two main types of video systems. Paralleling their audio counterparts, there are tape-based systems and disc-based systems. With the tape-based video systems you can, in your own home, both record and playback television programs and your own electronically produced home movies. The disc-based systems, by and large, are for replay of pre-recorded material only.

PORTABLE VIDEO SYSTEM

The Portable Video System *(tape-based)* is shown in Fig. 36.1. The *VC-20 camera* is capable of operating in almost any conditions and gives the user almost *total creative control* over the end result. The VR-10 *recorder* turns out to be the perfect partner for both the camera and tuner/timer, producing good pictures from either source. Features such as *insert edit* give this recorder the capability of producing truly professional home recordings. This facility, used with audio dubbing, gives the user the opportunity of achieving with one machine the kind of editing normally possible only in multi-machine studios.

The *tuner/timer* has all of the features of much larger domestic video recorders. Its electronic channel selection is an innovative idea and its programmable capabilities are more than adequate. *One touch recording* (OTR) will start recording a program at the touch of a button.

Fig. 36.1 Portable video system

As a complete unit, you have a very impressive combination which does far more than a home video recorder. The system offers the scope for truly creative movie making.

LASERVISION—VIDEO DISC SYSTEM

Laservision is without doubt, the most sophisticated video disc system. *Its future lies in the industrial/educational market* and not the mass consumer market. The disc itself is the same size as an audio LP but is *silver covered* and *reflective* like a mirror.

Stereo sound is available when the player is linked to a hi-fi and Laservision, Fig. 36.2, offers a full range of *special features* including slow motion, freeze frame and picture search.

The Philips discs also offer numerous *interactive possibilities* once they have been specially pre-programmed. For example, they could be used to test students by posing questions and then providing the correct answers.

Fig. 36.2 Laservision—Video disc system

INTERACTIVE VIDEO SYSTEMS

When video can be made to respond to the wishes and instructions of the operator, it is said to be *interactive*. Both video tapes and discs have a role to play in this new application of video, which promises to introduce new methods of learning.

The aim of interactive learning through video is to let the machine take over from the teacher. If it incorporates a programmed tape, a machine on which to play it, a monitor and a device, linked to a print-out, into which the student can punch responses, an interactive video system can check a student's progress and understanding step by step, and discover which areas of study demand the teacher's personal attention.

The next step in sophistication is to link the tape player of the system to a computer such as Apple II. *With the potential of a computer added to the inventory, the level of education can be made more flexible and more technical.*

Whenever interactive programs are used with VCR systems, they always meet the same drawback: the *access time.* This is the time it takes to spin the tape from the *end* of one section to the *start* of the next. The solution is to use video discs instead of tapes. *Because the way in which they are constructed and played, video discs allow virtually immediate access to any other part of the disc.* All the machine has to do is move its playback head a few centimetres and the next section of the recording appears on the screen. The player incorporates a microprocessor, so a hand held infra-red remote control pad can be used to interact with the disc. On advanced machines the sound can be set for stereo, bilingual or a choice of commentaries at *two levels of difficulty.*

Linked to a computer, interactive video discs are proving to have some fascinating examples. The American Heart Foundation, for example, has a disc on mouth-to-mouth resuscitation, in which the computer is linked to a dummy. As the first aider goes through the course, the program flashes back comments on progress such as ‘first breath too hard, third too weak’.

The *Video Responder,* Fig. 36.4, is designed to allow students to progress at their own pace. The tape that is inserted into the U-matic part of the system is programmed in a series of up to seven segments. At the end of each segment, the student is given a series of questions to answer relating to what has been shown. In writing the questions to accompany the video tape, the teacher may specify up to three attempts for each question, and make provision for branching into more detailed treatment of any topic.

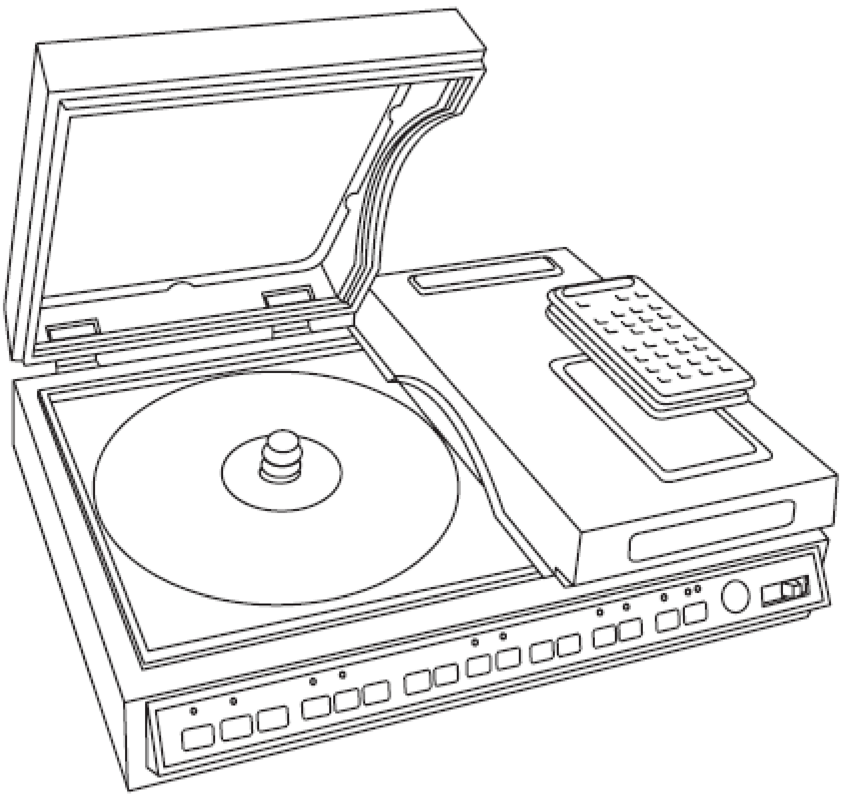


Fig. 36.3 Attached to a TV monitor, the disc vision player becomes an interactive video system.

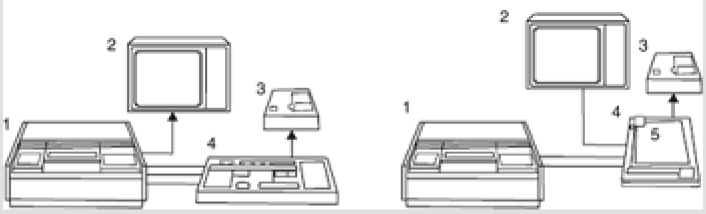


Fig. 36.4 Video responder system.

To make a program for the video responder system, the teacher first buys or records a video tape for play in the Randon Access U-matic, 1, connected to a TV monitor, 2. The Cue Programmer, 4, is then inserted into the circuit and used to write a series of questions to accompany each segment of the tape. The printer, 3 confirms the data entered at every stage.

In use during a training course, the student's rsesponder replaces the Cue Programmer, 4. After inserting the tape, and following the guidelines given in the work book, 5, the student presses appropriate buttons along the bottom edge of the responder. Answers may be recorded on the printer, 3, to give a permanent record of a student's progress at every stage.

Descriptive Questions

1. What are the different video systems?
2. Why disc based video systems are preferred?
3. Explain in detail an interactive video system. What are its practical applications?
4. What is the significance of a video responder?

Part II

ELECTRONIC GADGETS AND HOME APPLIANCES

UNIT – IV

TELECOMMUNICATION SYSTEMS

Telecommunications is defined as the technology concerned with communicating at a distance. The first requirement is for the original form of energy (human voice, music, or a telegraph signal) to be converted into electrical form to produce an electronic information signal. This is achieved by a suitable transducer, a device that converts energy from one form to another. In a telephone system, the transmitted signals are replicas of the speech waveforms. Digital signals consist of discrete pulses of voltage or current which represent the information to be processed. Digital voltages can vary only in discrete steps; normally only two voltage levels are used, so that two-state devices can be employed. A two-state device has two stable states; that is either ON or OFF. An electrical switch is a two-state device; it is either ON or OFF.

TELECOMMUNICATION SYSTEMS

In *a line communication system*, the electronic signal is passed to the destination by a wire or cable link, with the the energy travelling at a speed of upto 60% that of light, depending on the type of line. At the destination, a second transducer converts the electronic signal back to its original form. *Amplifiers* may be inserted to increase the power level of signals to compensate for losses encountered. This is shown in Fig. 37.1.

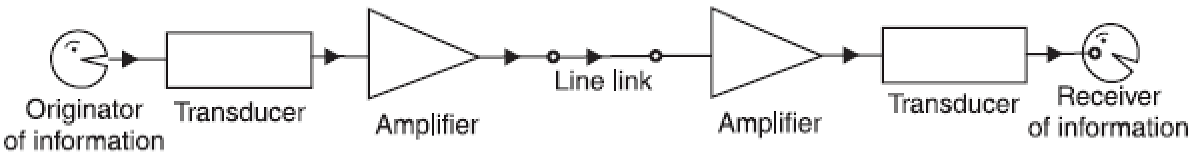


Fig. 37.1 Basic requirements of a one-way line communication channel

For a wireless communication system, a transmitter is required at the source to send the signal over the radio link, with the energy travelling at the speed of light, and a receiver is needed at the destination to recover the signal before applying it to a transducer.

In both of these system *interference* will be generated by electrical noise, and distortion of the electronic signal will occur for a number of reasons. These are *undesirable effects* and must be minimised in the system design.

Simple single-voice band signals are *unidirectional* (one-way only) and generally called channels; domestic radio and television broadcast are examples of such systems.

The telephone system, however, must be capable of conveying information in both directions. To do this the basic requirements must be duplicated in the opposite direction. A pair of complementary channels provide bidirectional communication, generally called a circuit.

Equipment is available which enables more than one voice channel to be carried on a pair of wires, a coaxial cable, a radio link or an optic fibre. Such *multichannel equipment* is called carrier or *multiplexing equipment.*

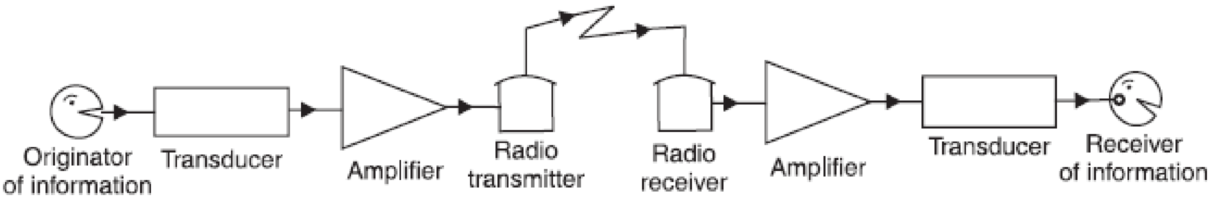


Fig. 37.2 Basic requirements of a one-way radio communication channel

Coaxial cable networks, with amplifying stations every few kilometres, now link together most of the cities in developed countries. These networks carry many thousand *multiplexed channels.*

Radio equipments operating at frequencies much higher than ordinary domestic radio sets are called *microwave links* are also able to carry thousands of multiplexed voice channels between terminals.

Optic fibres are a new and special form of transmission path in which energy representing up to two thousand voice channels can travel as pulses of light along a single glass or silica fibre comparable in diameter to a human hair. Optic fibre cable networks are now being installed in many countries making possible a huge expansion of telecommunication services.

LINE SYSTEM CHARACTERISTICS

The simplest form of two-wire line is produced by using bare conductors suspended on insulators at the top of poles (Fig. 37.3). The wires must not be allowed to touch each other; this would provide a short circuit and would interrupt communications.

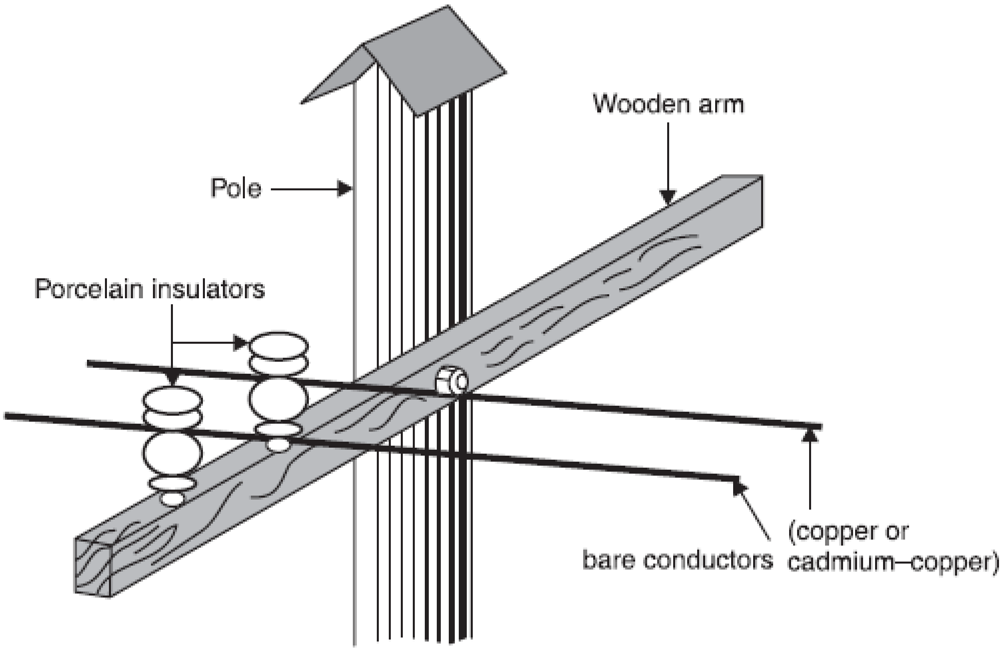


Fig. 37.3 Simple overhead two wire lines

Another type of two-wire line consists of conductors insulated from each other in a cable, which also has an outer cover of insulation (Fig. 37.4). This *outer sheath* used to be made of lead but various types of *plastic* are now commonly used, particularly PVC. The two insulated conductors in the cable are often *twisted* together along the length of the cable and are called a pair.

Many two-wire lines are often wanted between the same two places. These can most conveniently be provided by making a cable with a number of pairs of insulated wires inside it. Sometimes the wires are twisted together in pairs but sometimes they are provided in forms or *quads* as shown in Fig. 37.5.

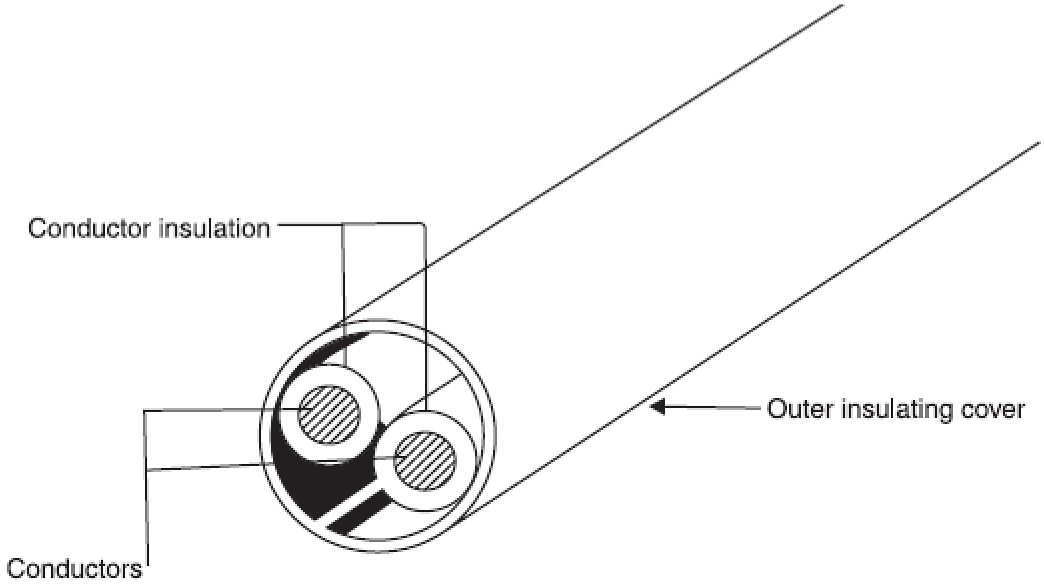


Fig. 37.4 Simple two-wire cable.

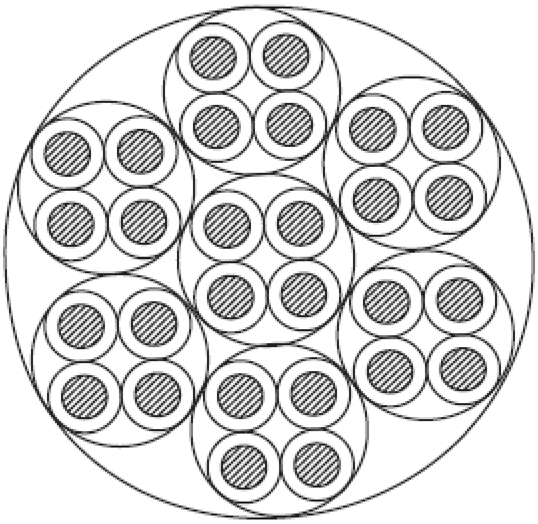


Fig. 37.5 Quad type cable

In order to identify the various wires each wire has a colouring on the insulating material around it, in accordance with a standard *colour code* for cable pair identification.

As the frequency of an alternating current is increased the current tends to flow along the outer skin of the conductor, and ordinary twin and quad type cables become inefficient. A special type of cable suitable for use at high-frequencies has therefore been developed. This has one of its conductors completely *surrounded* by the second one, in the form of a tube. This type of cable, called a *coaxial cable* is shown in Fig. 37.6. The two conductors can be insulated from each either by a solid insulant (*dielectric*) along the whole length of the cable or by insulating spacers fitted at regular intervals as supports for inner conductor. In this case the main insulation is the air between the two conductors.

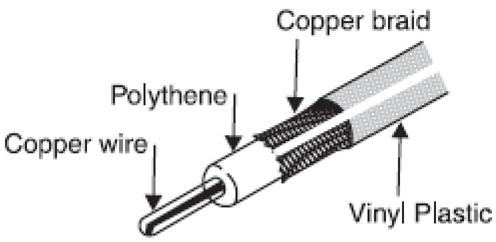


Fig. 37.6 Construction of a coaxial cable

Whatever the type of cable used, the conductors always have some opposition to current flow. This is called *resistance.* Furthermore no insulation material is perfect, so a very small *leakage current* will always flow between the two conductors, instead of all of it flowing along them to the distant end.

Also, the insulation between the conductors forms a *capacitance* which provides a conducting path between the conductors for alternating current. This capacitance has the ability to store electrical energy. The higher the frequency of the information signal, the more current travels across this capacitance path and the less reaches the distant end of the line.

When an electric current flow—along a wire, a magnetic field is established around the wire. Whenever the current in a wire *changes*, either by the switching on or off of a one way or direct current, or the *repetitive changes* of an alternating current, its accompanying magnetic field is made to change also, and energy is needed for these changes. This is called the *inductance* of the circuit.

Energy is used to make the current flow against the resistance along the conductors, and against the insulation resistance between the conductors. Energy is also used in charging and discharging the capacitance between the conductors. In multipair cables there is capacitive and inductive *coupling* between pairs also, so that some energy is passed from one pair to another. These losses further reduce the amount of energy that reaches the far end of the original pair and so contribute to the total loss.

In the case of an information signal, all this lost energy has to come from the signal source so that the energy available gradually decreases as the signal travels along the line. This loss of energy along the line is called *attenuation.* If the line is long and the attenuation is large, the received energy may be too weak to operate the receiving transducer unless some *corrective action* is taken.

RADIO SYSTEM CHARACTERISTICS

When a radio-frequency current flows into a *transmitting antenna* power is radiated in a number of directions in what is called an electromagnetic wave. These waves travel at the same speed as light and can be reflected and refracted just as light can be. Some antennas are designed to be highly directional, some are omnidirectional. The radiated energy will reach the *receiving station* by one or more of five different modes as shown in Fig. 37.7.

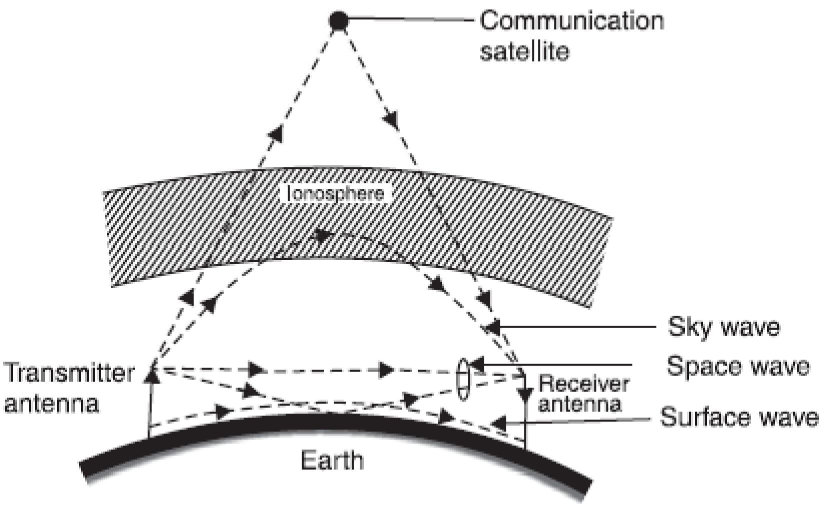


Fig. 37.7 Modes of propagation of radio waves

1. Surface wave;
2. Sky wave;
3. Space wave;
4. Via a satellite and
5. Scatter

1. **The surface wave** is supported at its lower edge by the surface of the earth and is able to follow the curvature of earth as it travels. The surface wave is used for world-wide communications in the low-frequency bands and for broadcasting in the MF band (see Table 37.1)
2. **The sky wave** is directed upwards from the earth into the ionosphere (100 km or more above ground level) whence, if certain conditions are satisfied, it will be returned to earth for reception at the required locality.
3. **The space wave** generally has two components, one of which travels in a very nearly straight line between the transmitting and receiving locations, and the other travels by means of a single reflection from the earth.
4. The fourth method is a technique that utilises the ability of a **communication satellite** orbiting the earth to receive a signal, amplify it, and then transmit it at a different frequency back towards the earth.
5. The fifth method, listed **scatter**, could be said to be the UHF/SHF equivalent of using skywave transmission for long distance HF radio links. The radio energy is directed towards part of the troposphere which forward scatters the signal towards the

receiver. The scattering region of the troposphere is about 10 km above ground level (Fig. 37.8).

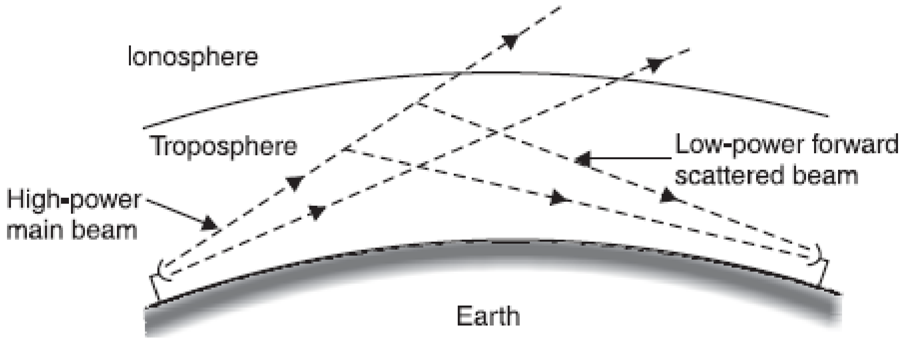


Fig. 37.8 Scatter propagation

The radio frequency spectrum has been subdivided into a number of frequency bands, these are given in Table 37.1.

Table 37.1 The Radio Frequency Spectrum

Frequency band	Classification	Abbreviation
Below 300 Hz	Extremely low	ELF
300 Hz-3 kHz	Infra low	ILF
3 kHz-30 kHz	Very low	VLF
30 kHz-300 kHz	Low	LF
300 kHz-3 MHz	Medium	MF
3 MHz-30 MHz	High	HF
30 MHz-300 MHz	Very high	VHF
300 MHz-3 GHz	Ultra high	UHF
3 GHz-30 GHz	Super high	SHF
30 GHz-300 GHz	Extremely high	EHF

300 GHz-3000 GHz	Tremendously high	THF

TELEPHONE RECEIVERS AND HANDSETS

Telephone receivers are *electromagnetic transducers* which convert telephone signals into sound pressure waves. They are made compact and of very rugged construction to withstand the abuse of telephone usage.

Figure 37.9 shows a compact version of the telephone receiver which was in use from the early 1920s until the late 1950s. Two small but very strong permanent bar magnets provide the magnetic bias field. This field is channeled to the diaphragm through the two formed soft iron pole pieces on which the coils are mounted as shown in Fig. 37.9. The whole magnetic structure is fastened to an aluminium frame which also supports the magnetic diaphragm, a perforated cover, and a dust cover membrane held in place by a formed aluminium collar. A fibre-board cover fastened to the back of the magnetic structure carries a contact button and a contact ring through which the electrical connections are made from spring contacts in the handset.

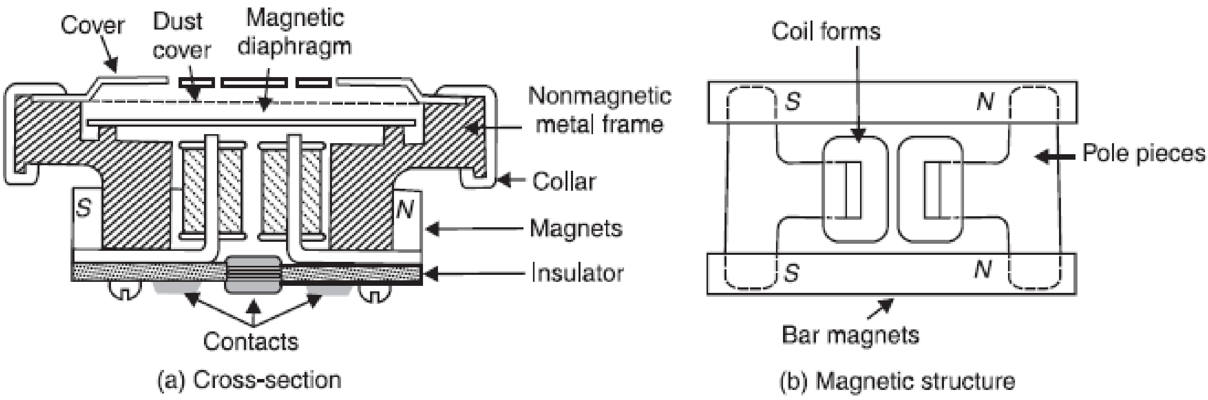


Fig. 37.9 Magnetic diaphragm type telephone receiver

The magnetic bias field is necessary because if it were not present, the diaphragm would vibrate at *twice* the frequency of the applied signal current. A magnetic diaphragm suspended above the poles of an electromagnet with no permanent magnetic field would experience attracting force when the current through the coils was zero. The diaphragm would experience maximum attraction force when the current reached the maximum value of either polarity. Thus the diaphragm would be pulled in twice during each cycle of the applied signal current. With a permanent magnetic field component present, the diaphragm always experiences a positive attraction force. With *maximum positive* signal, the force would be twice the zero signal force, and with maximum negative signal the force would approach a zero value. The diaphragm would thus oscillate around its bias position at the same frequency as the signal current.

The receiver described above has been almost entirely replaced by a different version, called *ring armature receiver* which was introduced in 1950s. Figure 37.10 shows a cross-section of such a receiver unit. In this unit the heavy magnetic diaphragm is replaced by a very lightweight dome shaped diaphragm made of aluminium which is fastened to a ring-shaped disk of magnetic material which forms the armature. The magnetic structure is formed by a cylindrical permanent magnet with an L-shaped cross-section and an L-shaped cylindrical soft iron pole piece attached to it. The driving coil is wound inside this magnetic structure, and the ring-shaped armature is suspended in the gap between the magnet and the pole piece supported by a non magnetic support ring. A bakelite backplate provides support for the screw terminals to the coil, and also provides an *acoustical resistance chamber* to damp oscillations of the diaphragm. In service a *bilateral varistor diode* is connected between the terminals to clip off or *limit* high amplitude voice pulses which may be generated by switching equipment. The acoustical chambers in front of and behind the diaphragm are carefully designed so that the mechanical resistance encountered by the diaphragm *closely matches* the electrical impedance of the telephone circuit. A perforated protection cover and a dust membrane are crimped around the top of the unit and the whole unit is mounted inside the ear piece of a handset. *Typical terminal impedances range from 100 to 2000 Ω.*

SIGNALLING

In telephony context *signalling means the passing of information and instructions from one point to another relevant to the setting up or supervision of a telephone call.* To *initiate* a call a telephone subscriber lifts the handset off its rest-goes off hook. This off hook state is a signal to the exchange to be ready to receive the number of the called subscriber. As soon as appropriate receiving equipment has been connected to the line, the exchange signals *dial tone* back to the calling subscriber who then dials the wanted number. On older exchanges, this information is passed by a rotary dial (Fig. 37.11) by a series of makes and breaks of the subscriber's loop interrupting current flow. On more modern exchanges, voice frequency musical tones are sent to the exchange as push buttons (Fig. 37.12) are pressed. These tones are usually called DTMF (Dual Tone Multi Frequency), because each time a button is pressed two tones are sent out of line simultaneously, one from a set of four *high*-frequencies, one from a set of four *low*-frequencies. The subscriber in due course then receives advice from the exchange about the status of the call, either a ringing signal (indicating that the wanted line is being rung), an engaged or busy tone signal (indicating that the wanted line is already busy on another call), an equipment busy tone signal (indicating congestion somewhere between the called exchange and the calling line) or some other specialised tone. These are signals and tones with which telephone subscribers themselves are concerned. Telephone signalling is however also concerned with the signalling of information between exchanges.

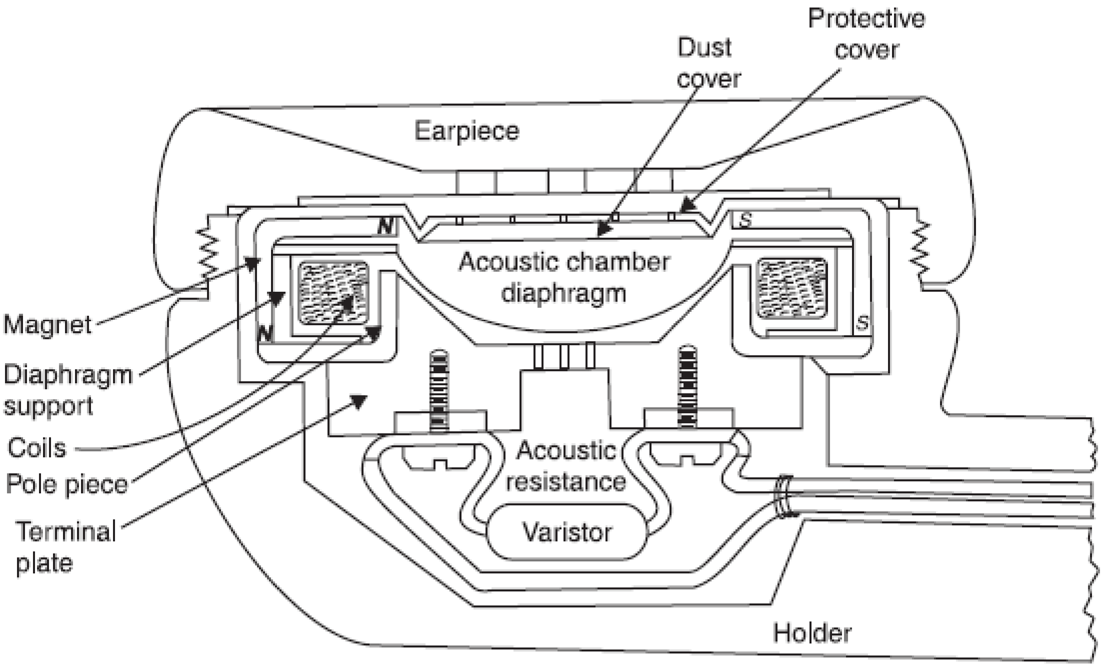


Fig. 37.10 Cross-section of a ring armature telephone receiver

Fig. 37.11 Rotary dial telephones

Fig. 37.12 Push button telephone

CCITT NO. 7

Until recently all such signalling was carried on or directly associated with the same speech path as was to be used for the call being established or supervised. Various terms are commonly used in connection with these *speech-path-associated signalling systems.*

1. *MF:* Multi-frequency, i.e. using voice frequency tones.
2. *MFC:* Multi-frequency compelled : this type of signal continues until the distant end acknowledges receipt and compels it to stop.
3. *1 VF:* One voice frequency : a single tone, sometimes pulsed in step with rotary dial impulses. *2600 Hz* is a common 1VF tone.
4. *2 VF:* Two voice frequencies : two tones, sometimes used together, sometimes separately
5. *In band:* A tone actually on the voice circuit itself, audible to any one using the circuit (and so cannot be used during conversation).
6. *Outband:* Signals directly associated with a voice circuit but either carried on separate wires or using a different frequency, just outside the commercial speech band of 300–3400 Hz. A frequency of 3825 Hz is often used for outband signalling.

All of these signalling systems have a number of *limitations :*

1. Relatively slow.
2. Limited information capacity.
3. Limited capability of conveying information which is not directly call related.
4. Inability of some systems to send detailed information back to the calling end.
5. Inability of some systems to provide sufficient information for accurate itemised call billing.

6. Systems tend to be designed for specific application conditions.
7. Systems tend to be expensive because each circuit has to be equipped independently; there are no sharing techniques and no economies of scale.

The increased use of computer controlled (or SPC, stored program control) exchanges has led to the introduction of a completely different signalling concept. Instead of signalling being carried on, or directly associated with, the voice channel carrying the conversation, *there is now a move towards signalling being concentrated onto fast data circuits between the processors of the SPC exchanges concerned leaving the voice circuits purely to carry voice signals*. Signalling for several hundred long-distance circuits can be carried by a single fast data system, and substantial economies result.

A signalling system of this type has now been standardised by the body responsible for drawing up specifications for international use; this is called CCITT Signalling System No. 7. CCITT means the International Consultative Committee for Telephony and Telegraphy. No. 7 signalling (Fig. 37.13), has not only been designed to control the setting up and supervision of telephone calls but of non-voice services also such as word processors, teletext machines etc. With Common Channel Signalling (CCS) systems, such as CCITT No. 7, signalling is performed in both directions, with one signalling channel in each direction. This type of signalling has several *attractive features :*

1. Signalling is completely separate from switching and speech transmission, and thus may evolve without the constraints normally associated with such factors.
2. Significantly faster than voice-band signalling.
3. Potential for a large number of signals.
4. Freedom to handle signals during speech.
5. Flexibility to change or add signals.
6. Potential for services such as network management, network maintenance, centralised call accounting.
7. Particularly economical for large speech circuit groups.
8. Economical also for small speech circuit groups due to the quasi-associated and dissociated signalling capabilities (sec Fig. 37.13).
9. Systems have been standardised for international use.
10. Can be used to control the setting up and supervision of non-voice services and so will be important for ISDN.

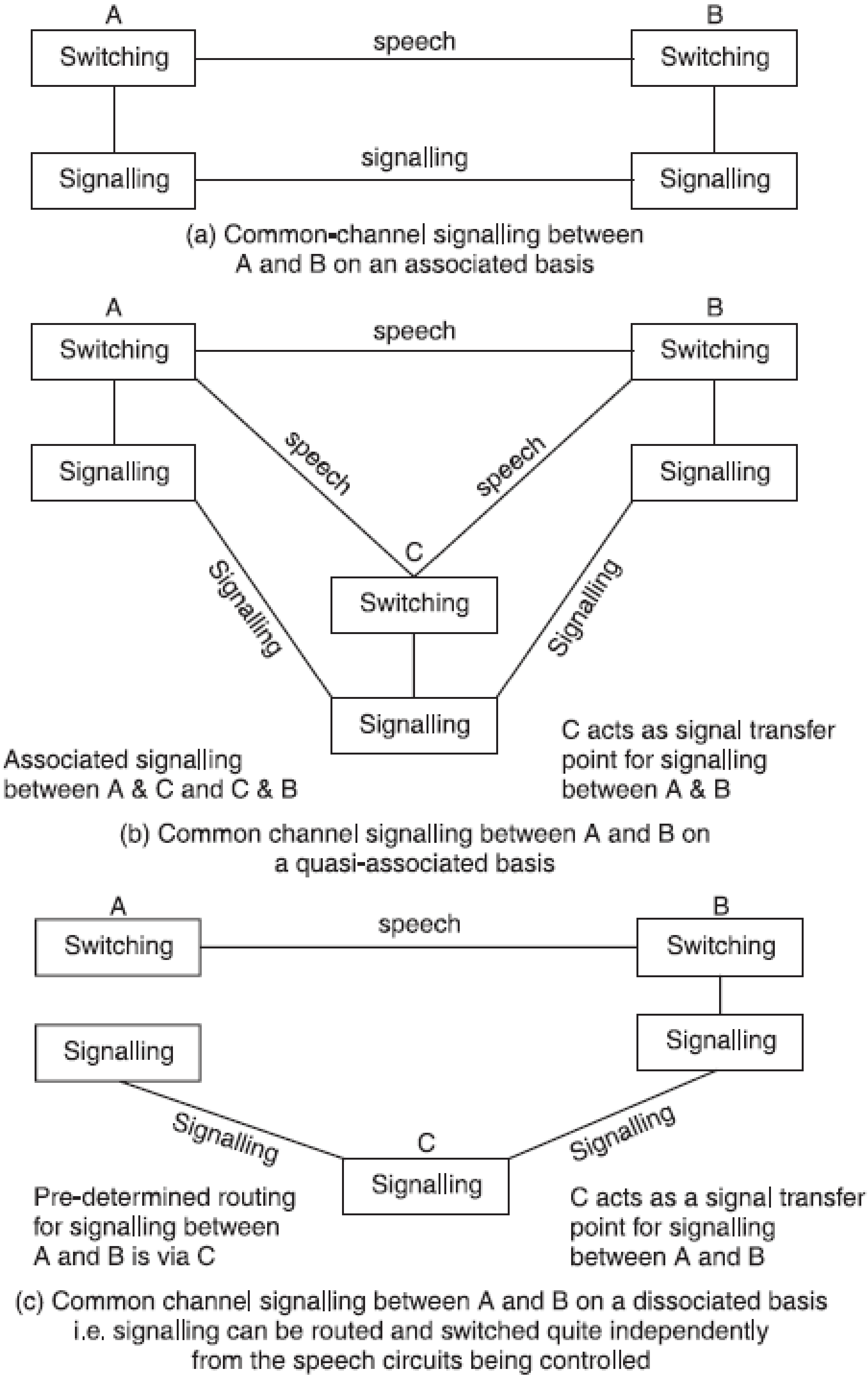


Fig. 37.13 Common channel signalling (such as CCITT No. 7)

A common channel signalling system providing signalling for many speech circuits must have a *much greater dependability* than on-speech-path signalling since random errors could disturb a large number of speech circuits. For this reason *provision must be made to detect and correct errors.* Additionally, automatic re-routing of signalling traffic to a good backup facility must occur in a situation of excessive error rate or on failure of a signalling link.

MODES OF OPERATION

It would be a good idea to look at some overall systems. Transmitters and receivers are *tied together* in a number of ways that describe the *kind of system* rather than the kind of transmitting format used. One of the first questions to come to mind is : In how many directions will the data be transmitted ? This can best be answered by defining the classifications of systems we use today (see Fig. 37.14).

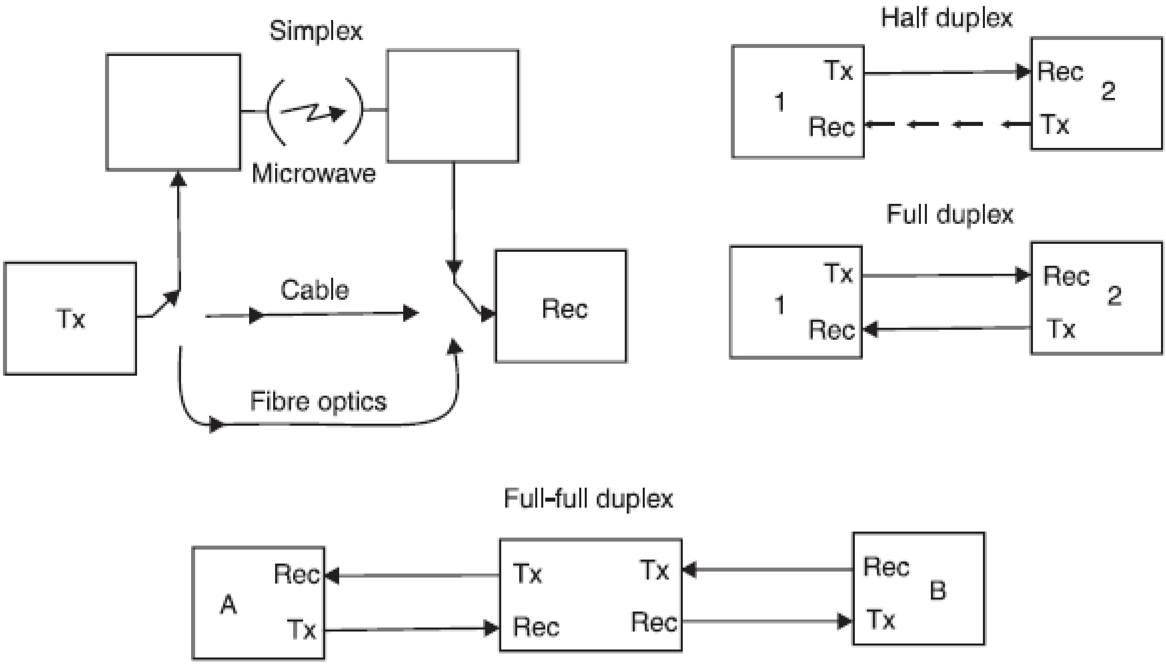


Fig. 37.14 Modes of transmission; simplex, half duplex, full duplex and full/full duplex

1. **Simplex (SPX) :** Information can be sent *in only one direction.* Commercial radio and television use this mode. They transmit, and you receive. You can never interact by the same communications link, regardless of how the signal reaches your receiver (RF transmission or cable).
2. **Half duplex (HDX) :** Information can be sent *in both directions, but not at the same time.* Citizens band or ham radio are prime examples. The operator must press a control to transmit and release the control to receive. In the transmit mode the receiver is disabled, and in the receiver mode the transmitter is disabled.
3. **Full duplex (FDX) :** Information can be *sent in both directions at the same time* without interference. The telephone is probably the best example. Both parties may talk at the same time and each is able to hear the other.
4. **Full/full duplex (F/FDX) :** This is the latest mode to be developed. It describes *a unit that is able to receive a message from one remote station while it is transmitting a different message to a third station.* Machines are able to carry on two or more conversations at the same time.

Although RF signals are used in these examples, most of the world's land communications are handled by the interconnecting network of telephone companies. The same definitions hold regardless of which route is taken. In RF transmission, when the system has *one carrier frequency* the land line has *one twisted wire pair*. When the RF system uses *two different frequencies*, the land line has two *twisted wire pairs* as would be required for FDX.

STATION INTERCONNECTIONS

The majority of digital traffic is processed through the *telephone networks* (see Fig. 37.15). An outgoing message is carried through twisted wire pairs to the *local branch exchange*, then to the receiver's local branch exchange, and finally to the subscriber's receiver. Long distance interconnections go from, the transmitter's local branch exchange to a toll exchange, from there to the destination toll exchange, then to the subscriber's branch exchange, and then to the subscriber's receiver.

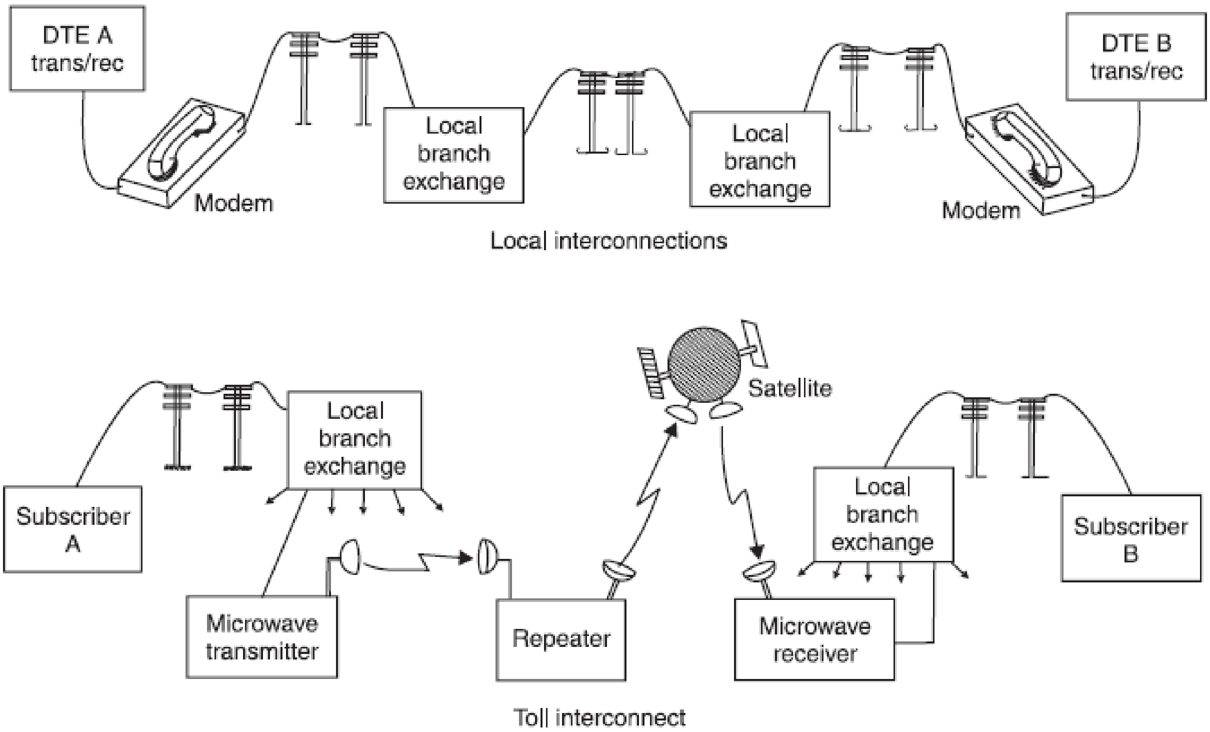


Fig. 37.15 Interconnections between subscribers: local and toll

Depending on the distance between the stations, the message may undergo any number of modulation and demodulation processes, including satellite links and fibre optic cabling, without the knowledge of the transmitter or receiver. *Since the telephone company systems are almost all automatic the equipment will pick the route of the interconnection that is open at the time of call origination.*

Such interconnection systems are used for what the phone company calls *dial uplines.* These are the same lines that are used for the home phone. The sender uses a standard telephone to dial the number of the receiver and then places the hand set of the phone into an *acoustical coupler* attached to the transmitting digital terminal. The equipment is compensated for the data speed the telephone company is able to handle.

When *high data rates* are a requirement, arrangements can be made for special routing, which the subscriber pays for on a leased basis. These lines are called *private or dedicated lines.* The subscriber may request a hard copper connection, which means that there will be no switches or transformers in the line between the transmitter and the receiver.

The above description applies to *serial code* transmission, using a two-wire or four-wire interconnection, but could also apply to *parallel code* transmission, where eight wires would be required for an eight-digit transmission. The tradeoffs between serial and parallel transmission in the distance between stations and daily leasing costs.

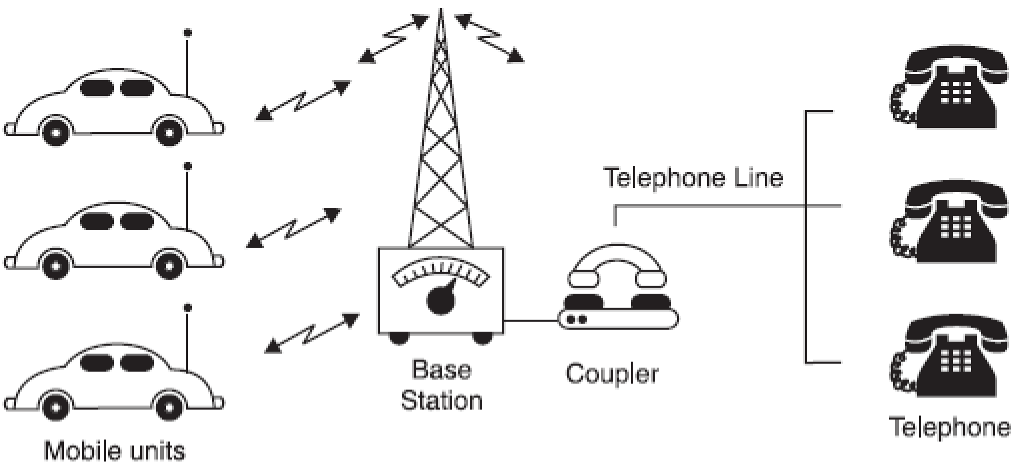


Fig. 37.16 An acoustic coupler used to interconnect mobile radio units to the telephone network

THE RS232C INTERCONNECTING CABLE

Because of the growing number of computer manufacturers, it became necessary to establish a fixed arrangement of interconnecting computers to match the connections of any unit with the connections of any other units. The solution became known as the 232C inter connecting cable (Fig. 37.17 and Table 37.2).

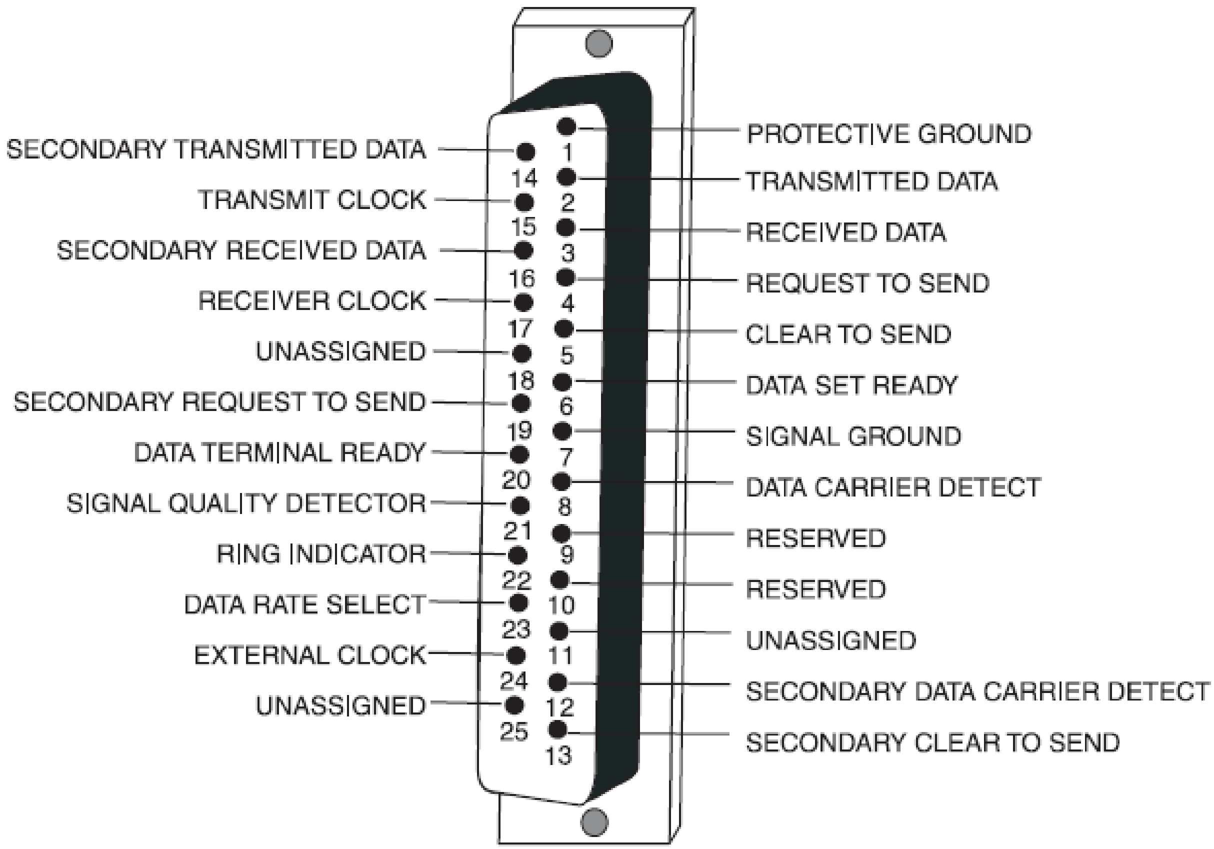


Fig. 37.17 The RS 232C interconnector

Table 37.2 The RS 232 C Interconnecting Cable

Pin	EIA Designation	Direction DTE DCE	Common Usage
1.	Protective ground		
2.	Transmit data	⟶	Send data (TD, SD)
3.	Receive data	⟵	(RD)
4.	Request to send	⟶	(RS, RTS)
5.	Clear to send	⟵	(CS, CTS)
6.	Data set ready	⟵	Modem ready (MR)
7.	Signal ground		
8.	Receive line signal detect	⟵	Carrier on detect (COD)
9.	Data set testing		+ V (+ 10 V DC)
10.	Data set testing		– V (– 10 V DC)
11.	Unassigned		
12.	Secondary (see 8)	⟵	Local mode (LM)
13.	Secondary (see 5)	⟵	
14.	Secondary (see 2)	⟶	New sync (NC)
15.	Transmission timing	⟵	Serial clock transmit (SCT)
16.	Secondary (see 3)	⟵	Divided clock transmit (DCT)
17.	Receiver timing	⟵	Serial clock receive (SCR)
18.	Unassigned		
19.	Secondary (see 4)	⟶	
20.	Data terminal ready	⟶	(DTR)
21.	Signal quality detector	⟵	(SQ)
22.	Ring indicator	⟵	(RI)
23.	Data rate select	⟶	(SS)
24.	Transmit signal timing	⟶	Serial clock transmit external (DSTE)
25.	Unassigned		

The RS 232 C generated by Electronic Industries Association, does not concern itself with the content or makeup of the serial data. Rather, it attempts to define *polarity* (or direction) and *level* of signals, as well as the *interfacing* of RS 232 C devices.

TTL levels are used extensively inside a computer. Most computers have a 5 V supply which powers all the TTL circuits. However, TTL is not intended to be run through wires for any great distance. Outside *electrical interference*, in particular, can mess up a TTL-level signal traveling within a wire cable. For example, let's say a serial printer in an office is 30–50 feet away from the computer. In this environment, interference from electrical cords and machines can scramble the digital data if TTL levels were used.

The makeup of a single digit byte, under the *RS 232 C specification* is shown in Fig. 37.18. A digital one is a minus voltage while a zero is a positive voltage. There is a *no man's* *land* between minus three volts and plus three volts, in other words, a *digital one* must be more negative than minus three volts, while a *digital zero* must be more positive than plus three volts. Thus, the polarity of an *RS 232 C bit is just the opposite of TTL.* Also the total amplitude is *more than* the five volt maximum found in TTL circuits. To meet the RS 232 C specification, the total amplitude must be at least six volts to go above and below the dead zone.

The purpose of this six-volt zone is *noise protection.* Static which might be coupled into the RS 232 C cable must exceed at least three volts before it will foul up the digital data. Such noise levels are not very likely even in the presence of strong electrical noise.

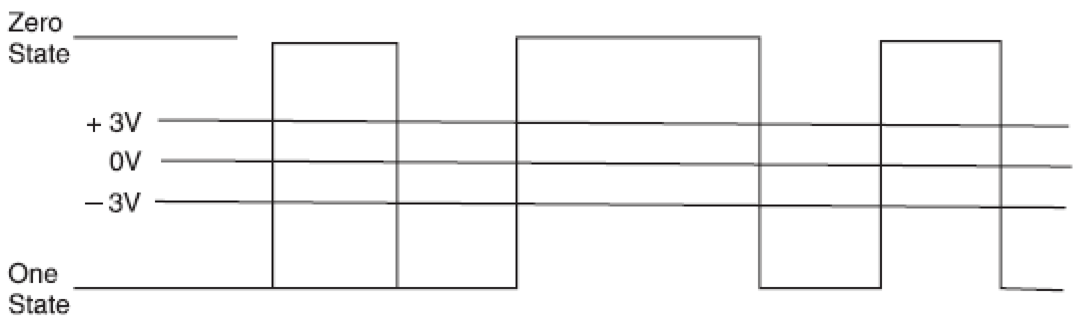


Fig. 37.18 The RS 232C specification includes a dead zone. The signal must be above or below these limits

EXERCISES

- **Copy**
- **Add Highlight**
- **Add Note**

Descriptive Questions

1. Differentiate between a wireless and a line communication system.
2. What are the characteristics of a wireless communication system?
3. What are the characteristics of a line communication system?
4. Briefly explain the following terms:
    a. Interference
    b. Unidirectional and bidirectional channels
    c. Multiplexed channels
    d. Microwave links
    e. Coaxial cable
    f. Attenuation
    g. Modes of propagation of radio waves
    h. Scatter propagation
5. Give examples of electromagnetic transducers. With the help of a suitable diagram explain the working of a ring-armature receiver.
6. What are the attractive features of common channel signalling (CCS)?

7. What are the classes of communication systems? Give examples.
8. How telephone networks are interconnected?
9. Explain the following terms:
    a. Local
    b. Toll
    c. Acoustic coupler
    d. Modem
10. Briefly explain the RS 232 C interconnecting standard. Why is it required?

Fill in the Blanks

1. Interference and distortion are _______________ effects.
2. Coaxial cable networks carry many thousand _______________ channels.
3. The two insulated conductors in the cable are often twisted together along the length of the cable, and are called a _______________ .
4. There is a standard colour code for cable pair _______________
5. In the case of an information signal, all the lost energy has to come from the _______________
6. The surface wave is able to follow the _______________ of the earth as it travels.
7. The skywave is directed upward into the _______________ .
8. The scattering region of the _______________ is about 10 km above ground level.
9. Typical terminal impedances range from _______________ to _______________ .
10. To initiate a call a telephone subscriber goes _______________ .
11. In DTMF each time a button is pressed _______________ are sent out to the line simultaneously.
12. In a CCS system provision must be made to _______________ errors.
13. In a simplex system, information can be sent in _______________ .
14. In a half duplex system, information can be sent in both directions but _______________ .
15. In a full duplex system, information can be sent in _______________ at the same time.

ANSWERS

Fill in the Blanks

1.	undesirable
2.	multiplexed
3.	pair
4.	identification
5.	signal source
6.	curvature
7.	ionosphere
8.	troposphere
9.	100Ω, 2000Ω
10.	off hook
11.	two tones
12.	detect and correct

13. only one direction
14. not at the same time
15. both directions

CHAPTER 38

SWITCHING SYSTEMS

Telephony was invented in the 1870s; all the early exchanges used human operators to establish and supervise calls. As networks grew it became impractical to continue to use people to set up telephone calls. It has indeed been calculated that to carry today's telephone traffic using 19th century practices would need more than half the total population of all major cities to be employed as telephone operators.

Modern telephones are a far cry from Bell's original model and even from the ones available a couple of decades ago. They vary in shape and features. They have push buttons instead of dials and memories to store telephone numbers. Some are cordless—the handset is not attached to the base unit, but is linked to it by short range radio. However, the principles of telephony have remained the same from the beginning.

SWITCHING SYSTEM PRINCIPLES

A switching system of some sort is needed to *enable* any terminal (e.g. a telephone, a teleprinter, a facsimile unit) to *pass* information to any other terminal, as *selected* by the calling customer.

If the network is small, direct links, (Fig. 38.1) can be provided between each possible pair of terminals and a simple *selecting switch* installed at each terminal. If there are five terminals, each must be able to access four links, so if there are N terminals there must be a total of ½N (N–1) links.

A slightly different approach would be *to have one link permanently connected to each terminal*, always used for calls to that particular terminal (Fig. 38.2). Again, each terminal would need a *selection switch* to choose the distant end wanted for a particular conversation, but the number of links is reduced from 10 to 5 for a 5-terminal system, and to N links for N terminals.

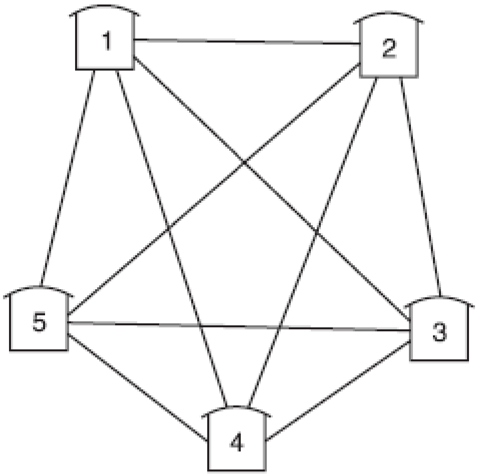


Fig. 38.1 Full interconnection

As number of terminals and distances increase, this type of arrangement becomes impossibly expensive with today's technologies. A variant of this is however in wide use in radio telephone networks : *all terminals use a single common channel to give the instructions for setting up each call* (Fig. 38.3). The terminals concerned then both switch to the allocated link or channel for their conversation. The number of links may be *reduced substantially* by this method; enough need to be provided only to carry the traffic generated by the system. But each terminal still needs access to several channels and must have its own *selection switch.*

So far as *telephone networks* are concerned it is at present more economical to perform all switching functions at *central points* i.e. not to make each terminal do its own switching. This means the provision of only one circuit from the nearest switching point or *exchange* to each subscriber's terminal (see Fig. 38.4). As networks grow and expand into other areas more *switching points* are provided, with special circuits between such points to carry traffic between the areas concerned. The whole network has to be designed to strict *transmission parameters* so as to ensure that any subscriber on any exchange can converse satisfactorily with all other subscribers, anywhere in the world. *Call accounting equipment* also has to be provided so that appropriate charges can be levied for all calls made.

The first telephone exchanges were *manual* and all calls were established by operators. *Automatic* exchanges using electromechanical relays and switches were however developed very rapidly. *Computer-controlled* exchanges, with no moving parts at all, are now beginning to become common. Most countries in the world now have automatic telephone systems fully interconnected with the rest of the world's systems.

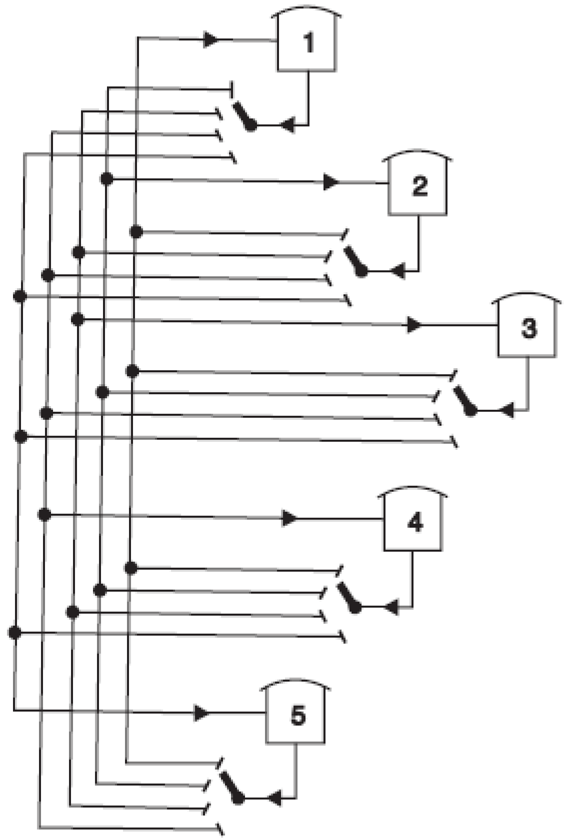


Fig. 38.2 One link per terminal method

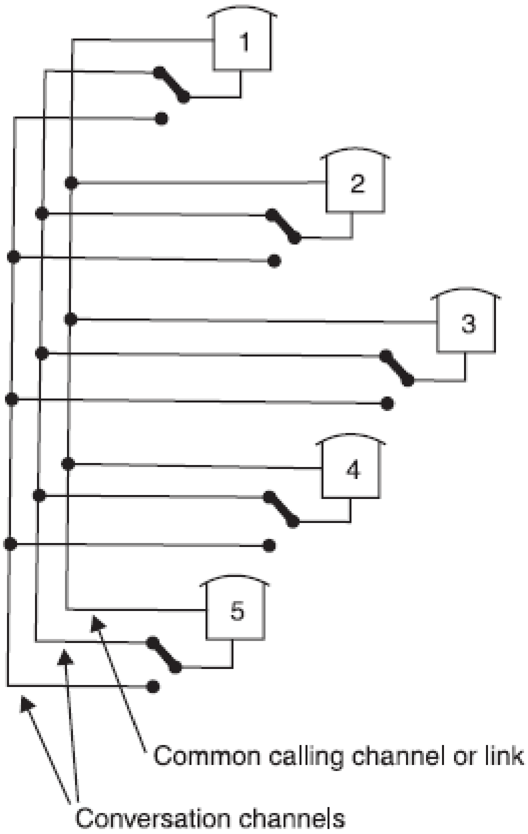


Fig. 38.3 Use of a calling channel

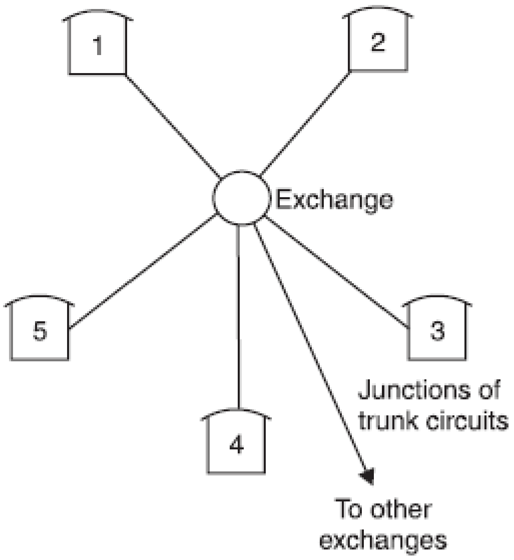


Fig. 38.4 Use of a switching centre; the telephone exchange

UNISELECTORS

For automatic operation the first requirement is a way of indicating the *telephone number* of the customer to whom you wish to speak. The *rotary dial* with ten finger holes is now nearly a century old in basic concept but is still in wide use. Contacts within the dial make and break an electrical circuit which *interrupts* current flowing, from a battery in the exchange, through the *loop* made by the line to the customer's premises and through the phone itself. If for example, you dial 6 the dial breaks circuit 6 times with each break lasting a pre-determined time, usually about 1/20th of a second. Relays in the exchange respond to these break signals.

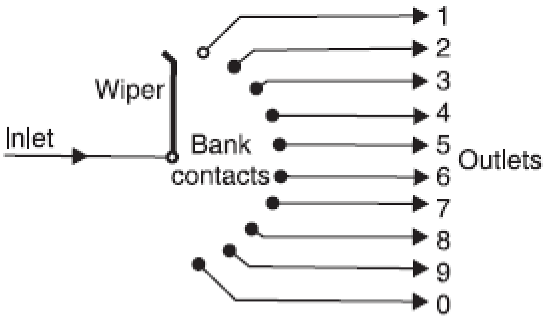


Fig. 38.5 Principle of switching by electromechanical uniselector of one-from-ten outlets.

The *step-by-step principle* was the first automatic system to become practical for public telephone exchanges; the selection of a particular line is based on a *one-from-ten selection process.* Fig. 38.5 shows a simple switch that has ten contacts arranged around a semicircular arc or bank, with a rotating contact arm or wiper that can be made to connect the inlet to any one of the ten bank contact outlets as required. The *wiper* is rotated by a simple electromagnet driving a suitable mechanism, so the arrangement is called an *electromechanical switch.* The wiper rotates in one plane only, so this type of electromechanical switch is called a *uniselector.* Clearly, the *inlet* can be connected to any one of the ten outlets, but the *outlets* are numbered from 1 to 0, which is normal practice in the step-by-step switching system.

This principle can be extended to enable the inlet to be connected to any one from 100 outlets by connecting each of the ten outlets of the first uniselector to the inlet of another uniselector as shown in Fig. 38.6. The switching of the inlet to any one of the 100 outlets (numbered 11 to 00) is done in two steps, the *first digit* being selected on the first uniselector, and the *second digit* being selected on the second uniselector.

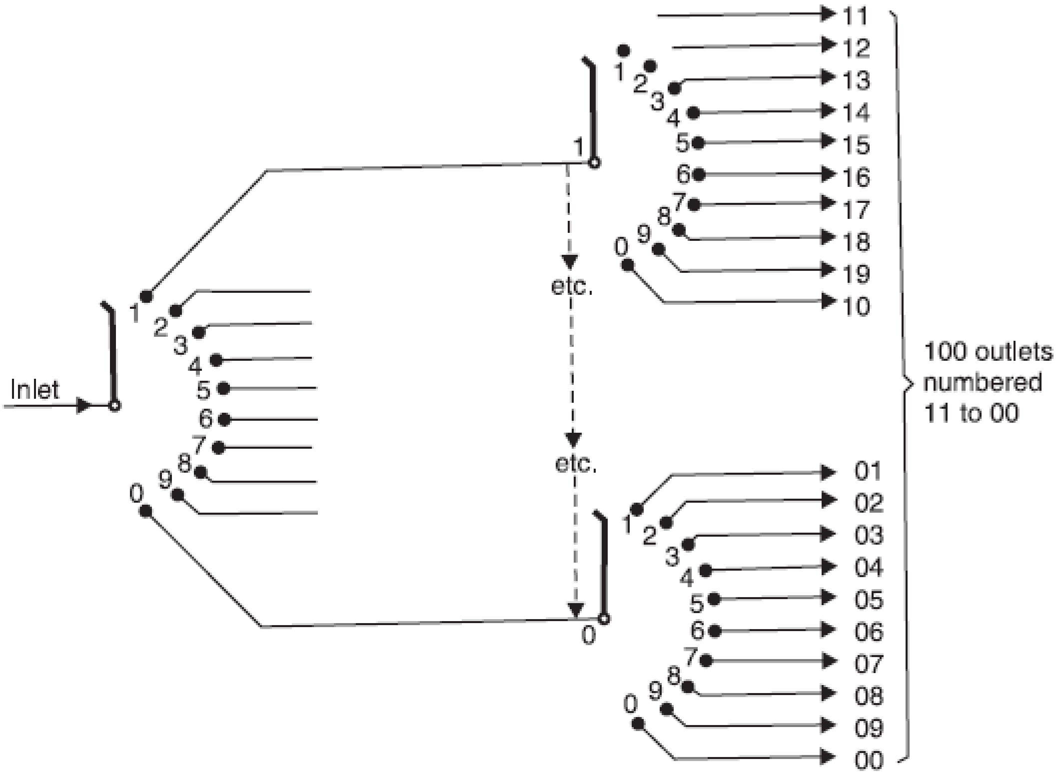


Fig. 38.6 Simple step-by-step selection of one-from-a-hundred outlets

If each of these 100 outlets is now connected to another uniselector, the inlet can then be connected to any one from 1000 outlets, numbered from 111 to 000, with the digits being selected one at a time on *three successive switching stages.* This arrangement can theoretically be extended to accomodate any number of digits in a particular *numbering scheme.*

TWO-MOTION SELECTORS

The same sort of numbering scheme can be provided on a step-by-step basis, by a different type of electromechanical switch called a *two-motion selector*. The principle is illustrated in Figs. 38.7 (a) and (b). The bank of *fixed contacts* now contains 10 semi-circular arcs, each having 10 contacts, and arranged above each other. The *moving contact* or wiper can be connected to any one of the 100 bank contacts by first *moving vertically* to the appropriate level, and then rotating horizontally to a particular contact on that level. The 100 outlets are numbered from 11 to 00. The diagram symbol used to illustrate the 100-outlet two-motion selector is shown in Fig. 38.7(c) .

As with the uniselector arrangements, the two-motion selector system can be *extended* to give access to any number of oulets by adding an extra switching stage for each extra digit required in the numbering scheme. A three digit numbering scheme from 111 to 000 outlets is illustrated in Fig. 38.8 with the one-from-a-hundred selector preceded by a one-froms-ten selector.

In Fig. 38.8, the *first digit* of the three digit numbering scheme raises the wiper of the first two-motion selector to the appropiate vertical level the selector, then *automatically searches for a* *free outlet on that level* to the next selector which caters for the *last two digits* of the three digit number, as shown in Fig. 38.7(a).

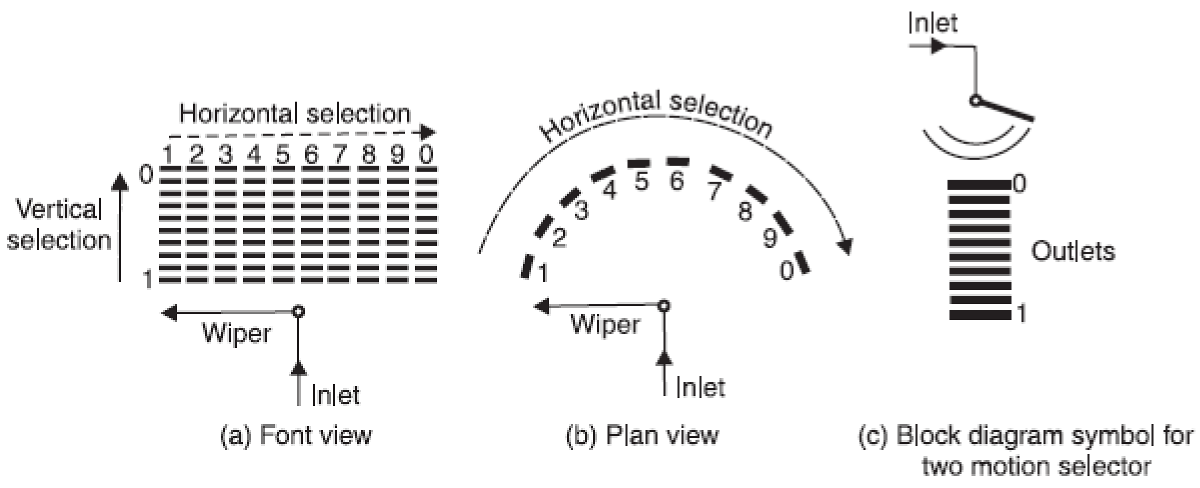


Fig. 38.7 Principle of one-from-a hundred selection by two-motion selection

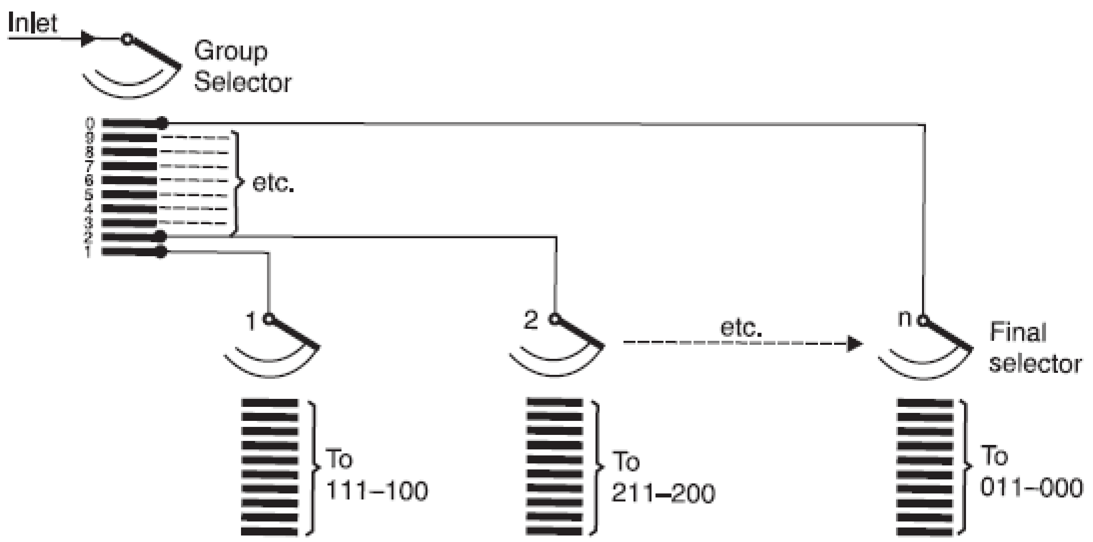


Fig. 38.8 Theoretical selection of one-from-a-thousand by two stage step-by-step switching

FOUR DIGIT STEP-BY-STEP AUTOMATIC EXCHANGE

In order to provide access to 10000 lines, a further stage of group selectors is added before the final selectors. Fig. 38.9 illustrates how a calling subscriber can be connected to other subscribers in an exchange having a *four digit numbering scheme*. Theoretically, a four digit numbering scheme can accomodate 10000 subscribers, but it is necessary also to provide junctions to other exchanges, lines to the operator and other enquiry services, and so on. This means that the *capacity* of an exchange of this type would *in practice* be about 6000 subscribers instead of the *theoretical* 10000.

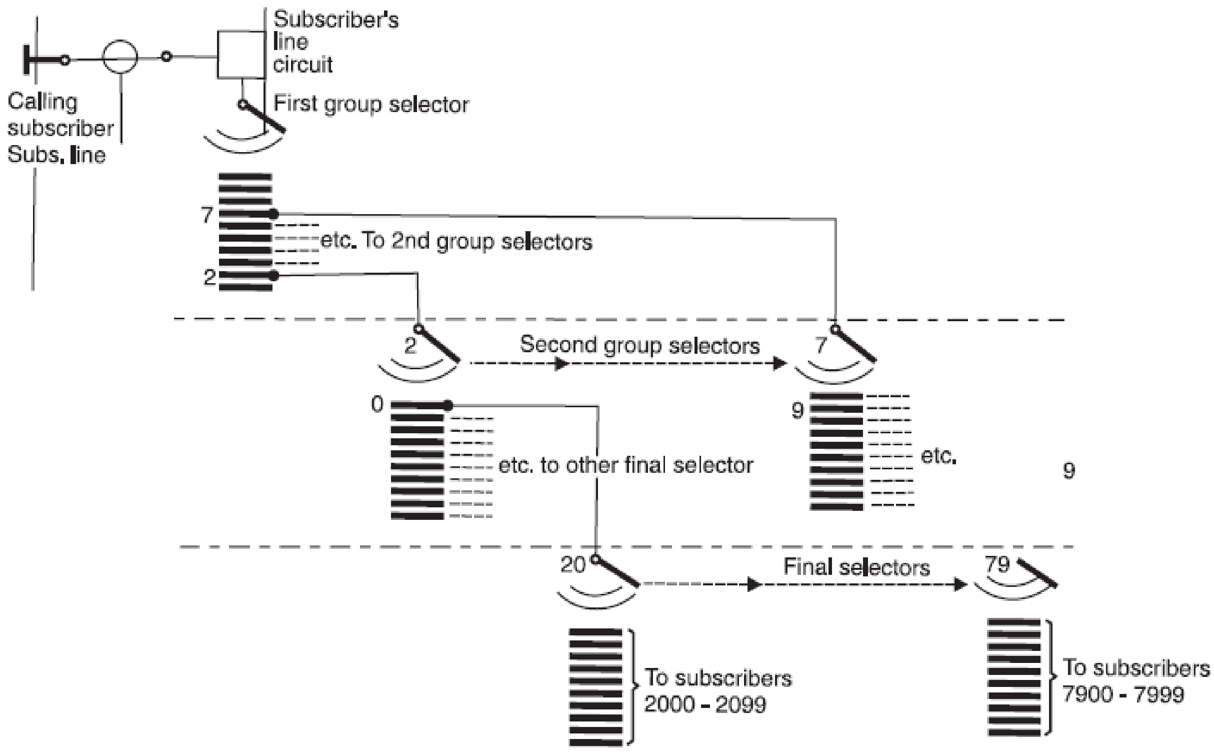


Fig. 38.9 Simple trunking diagram of four digit step-by-step automatic exchange

One of the basic features of step-by-step exchanges is that each selector or switch is conrolled by a group of electro-magnetic relays which is in effect a *small brain* just sufficient to act on the digit it receives and to route the call to another selector which acts on the next digit, and so on. Each digit dialled takes the caller one step nearer to the called number. In some countries' step-by-step exchanges and selectors are called *Strowger exchanges* and *Strowger switches.*

Most of the group of relays associated wih each selector in a step-by-step system are only used while the call is being set up so, as soon as the appropriate digit has been received and the selector stepped to the particular number dialled, most of these relays are *idle.* In a main exchange there are likely to be several thousands of these complex selectors, many of them will only be brought into use during busy periods and even the busiest selectors only use all their *brains* for a second or two every few minutes. It follows, therefore, that *a large amount of expensive equipment in such step-by-step exchanges remains idle for most of the time*.

Step-by-step selectors are robust and the principles are easy to follow, so *fault finding is usually fairly straight forward and faults can be speedily rectified.* The selectors do however sometimes introduce an unacceptable level of *noise* into conversations; they shudder quite violently while the calls are being set up and while they are releasing at the end of each call these movements affect electrical resistance at contact points and so produce noise in circuits.

Selectors have a great many moving parts which means that regular lubrication, with cleaning of relay contacts and readjustment of switches from time to time is absolutely essenial if good service is to be maintained. The *high cost of maintenance* is one of the main reasons for the fact that electro-mechanical step-by-step exchanges are, in most counries, now being replaced by exchanges which use low-maintenance-cost electronic techniques.

REED RELAY AND CROSSBAR EXCHANGES

A reed relay is a device based on the fact that an electric current passing through a coil of wire produces an electromagnet, with the ends of the coil having *opposite* magnetic polarities, as shown in Fig. 38.10.

These two strips can be used to form a switch in another electrical circuit. The two strips are placed inside a glass envelope containing an inert gas, and the *overlapping portions* are coated with gold to give a good reliable electrical contact. The whole assembly contained by the glass envelope is called a *reed insert*, since it is placed inside the electromagnet coil.

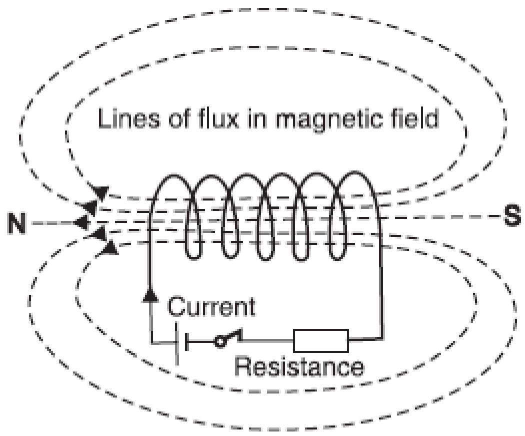


Fig. 38.10 Coil of wire as a simple electromagnet

A typical reed relay has four of these reed inserts placed inside the electromagnetic coil, each of which can be used to switch a separate electrical circuit. Coils are usually then arranged in a *matrix formation* so that the contacts in any particular reed relay in the matrix may be operated under the control of pulses of current through winding coils.

Selectors in *crossbar exchanges* have horizontal and vertical bars operated by electromagnetic relay coils, so that, with a *crossbar switch* also, the contacts at a particular point in a matrix may be operated under the control of these relays.

Crossbar switches and reed relays are both used in telephone exchanges. The basic concept is however quite different from that of step-by-step exchanges. Instead of each switch or selector having its own little *distributed* brain, there is a *central* brain which controls all switches (see Fig. 38.12).

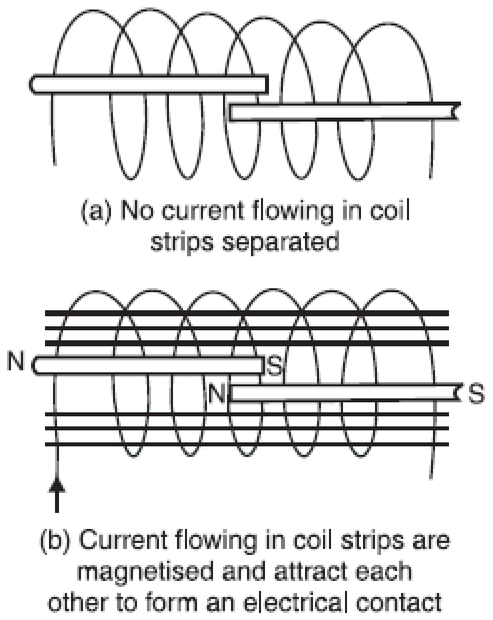


Fig. 38.11 Principle of operation of a reed relay

This central brain or *register marker* is just like a computer; it registers the number dialled, it checks that the calling number is permitted to make the call, and tests to see if the called number is engaged. Exchanges using this centralised control function are called *common control exchanges.* If the called number is free, this common control equipment chooses a pass through the exchange to join together the calling line to the called line, and issues instructions to all the crossbar switches or reed relays concerned to operate in such a way that the two lines are connected together. All this happens very rapidly—*calls do not have to be switched through the exchange one digit at a time in step with subscriber's dial.* This increased speed of operation means that it becomes attractive to replace ordinary rotary dials by *push buttons*; these allow calls to be set up very rapidly indeed.

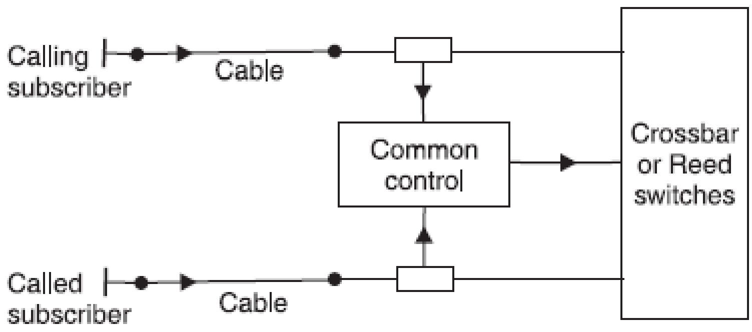


Fig. 38.12 Principle of reed relay and crossbar exchange

As there is less mechanical complexity in a crossbar or reed relay system, it is *more reliable* than step-by-step. Also the path through the exchange is such that crossbar switches and reeds *introduce less noise* into a telephone conversation than step-by-step switches.

Crossbar switches and reed relays have no moving parts demanding regular maintenance and adjustment, so *very little routine maintenance is needed and fewer exchange men are required.* It is this reduction of maintenance staff which usually results in large common control exchanges being more economic than similar size Strowger step-by-step exchanges.

TRAFFIC HANDLING CAPACITY

Both crossbar and reed relay switching depend on the opeation of a *switching matrix*, the principle of which can be explained by considering the circuits which are to be connected together as being arranged at right angles to each other in horizontal and vertical lines. These lines represent inlets and outlets of the switch. This idea is illustrated in Fig. 38.13. The intersections between horizontal and vertical lines are called *crosspoints.* At each crosspoint some form of switch contact is needed to complete the connection between horizontal and vertical lines. This is shown in Fig. 38.14. Any of the four inlets can be connected to any of the four outlets by closing the *appropriate* switch contacts. For example, inlet 1 can be connected to outlet 2 by closing contact B and inlet 4 can be connected to outlet 3 by closing contact R.

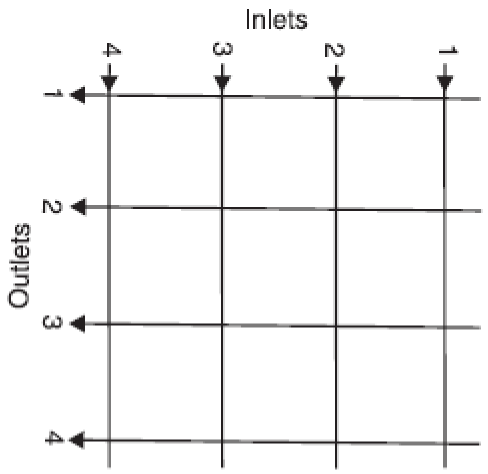


Fig. 38.13 Simple 4 × 4 switching matrix

The number of crosspoints in any matrix switch can be calculated by multiplying the number of inlets by the number of outlets (Fig. 38.15). If there are n inlets and m outlets, then the number of crosspoints is (n × m). If n is larger than m *(more inlets than outlets)*, then not all the inlets can be connected to a different outlet. When all the outlets have been taken, there will be some inlets still not in use. If m is larger than n *(more outlets than inlets)*, then when all inlets are each connected to an outlet, there will be some outlets still not in use. So the maximum number of simultaneous connections that can be carried by a matrix switch is given by *whichever of the number of inlets or outlets is smaller.*

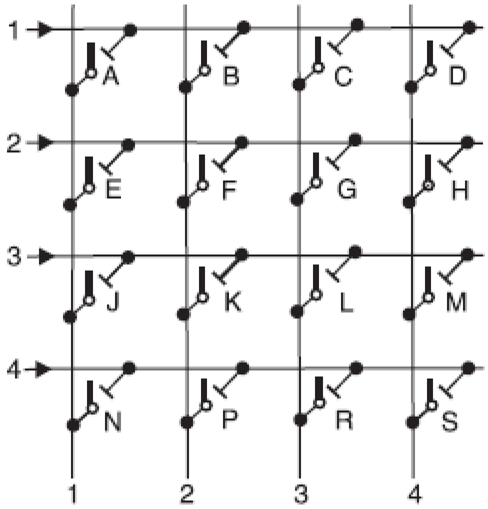


Fig. 38.14 Principles of switching by a 4 × 4 matrix switch

Efficiency in use of crosspoints

$$= \frac{\text{Maximum number of crosspoints that can be used simultaneously}}{\text{Total number of crosspoints in one matrix}} \times 100 \qquad \ldots(38.1)$$

For a large matrix this efficiency is necessarily very low, e.g. a 15 × 15 matrix with 225 crosspoints is only able to use 15 of these at any one time, giving only 6.7% efficiency.

Efficiency can be improved by using smaller matrix switches linked together. Most crossbar and reed relay exchanges are designed on this basis, with a series of interconnected switches. This can, in some circumstances, lead to link congestion or internal blocking. Careful design of the exchange is needed in order to maximise its traffic handling capacity while minimising equipment quantities, and therefore costs.

STORED PROGRAM CONTROL

Conrol in telephone exchanges developed from *individual control* of each switch on step-by-step exchanges to the use of a small number of complex centralised units in so called *common control exchanges*, which were mainly crossbar or reed. Common control units were originally completely electromechanical, using the same basic type of relay which had been used for many years in earlier exchanges. *Electronic common control* had to wait for the popularisation of the transistor and the printed circuit board before it became truly economic.

The use of computers to conrol telephone exchange switching is called *Stored Program Control* or SPC. This has been defined as *the control of an automatic switching arrangement in which call processing is determined by a program stored in an alterable memory.*

The basic function of an SPC system (Fig. 38.16) is *to control line originations and terminations and to provide trunk routing to other central or tandem offices*. The SPC system also provided control of special features and functions of a central office, identified here as *ancilliary control.* The intelligence of an SPC system resided in one *processor*, and all peripherals were controlled by this single processor. These processors were duplicated for reliability. Control of *maintenance functions* of the modern digital switching system also evolved from earlier SPC systems. These systems depended heavily on a single processor to conduct all maintenance functions of the switch. *Most of the modern digital switching systems employ a separate processor for maintenance functions.*

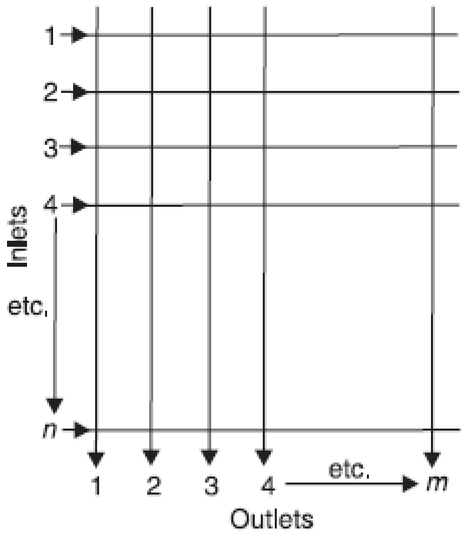


Fig. 38.15 Number of crosspoints in a matrix switch

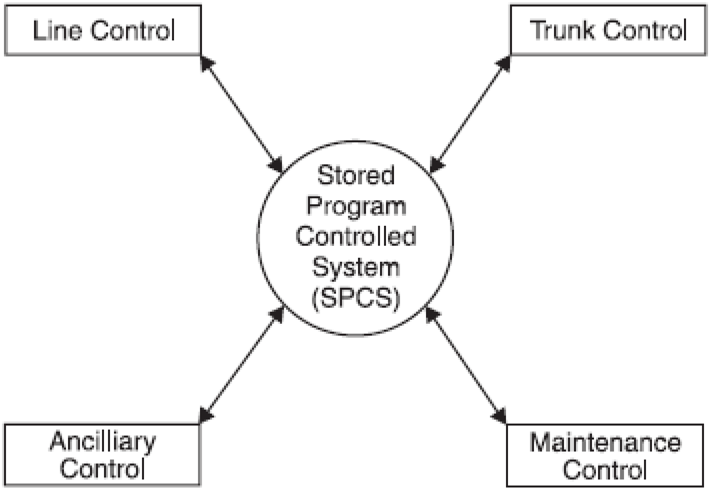


Fig. 38.16 Basic control structure of a central office

PBX SWITCHING

The acronym *PBX* (or *PABX* in Europe) stands for private auomatic branch exchange, a device for *switching* telephone calls within a building such as an office block or factory. The fundamental principles of PBX switching are the same as those found in a much larger central office switch, which is used in the public network to switch telephone traffic from the caller to the person who is being called. *The PBX is essentially a connecting device.* When a call comes into the PBX from an outside trunk line, it is either routed directly to telephone extension or to a console with a human attendant. The console rings and is answered by an attendant, who then connects the incoming call to the required extension by dialing the extension number required and, when this is answered, hanging up so that both parties can be connected.

Before PBXs were automated and computerised, the attendant *manually* connected both the caller and the person being called by physically inserting a cord which contained the incomming call, into a board full of sockets each of which was a designated telephone extension. A modern digital PBX attendant's console looks much different. *Indicator lights* on the console display which of the incoming lines is active, which callers are on hold, a display showing which extension the incoming call is being connected to and so on.

Since the early 1970s, PBX technology has moved through a number of distinct technological generations. Early PBXs were entirely *electromechanical*, using either *Strowger* or crossbar switching. With the introduction of computers, *SPC exchanges* replaced electromechanical switching. These PBXs, however, still operated analog switching and transmission, and far from all the PBX functions were compuerised. The next step was to computerise digital switches fully, with all the moving parts replaced by *microcircuitry*. Future PBX generations will continue the same digital theme, moving toward broadband switching and convergence with local area networks.

The fundamental elements of a digital PBX are shown in Fig. 38.17. All incoming lines enter some sort of *switching matrix* (1). Depending on the number of extensions connected to the PBX, there can be hundreds or even thosands of different switching paths to which the incoming call may be connected. The paths within the switching matrix are controlled by a central processor *(computer* (2)). Each extension line must be constantly *scanned for activity* (3) and the processor alerted, when activity occurs. The central processor must be informed when a user wishes to place a call as the telephone is picked up. When the number has been dialed, the processor must be able to read the required number and make the connection internally or if it is a call outside PBX, the processor must allocate it an outside line. Similarly, when a caller replaces the telephone handset after the call is finished, the processor needs to be informed that this extension is now *free* to accept further calls. There are also various *signals* which must be sent down the line, such as busy signals, *signals* to make the phone ring, on hook/off hook, dialing and so on. The *signaling equipment* (4) is controlled as a subset of the central processor. Finally, information about each call, its duration, the number called, and so on, needs to be gathered for *accounting and analysis* (5).

A *ditigal PBX* can handle both voice and data traffic and be *programmed* to support many different functions and applications. The *capacity* of a PBX to process calls and undertake the other functions that may be required of it is a function of the capability of the central processing unit. Microprocessor technology expands ever upward, and eventually it will be possible to put almost, all the functions of a PBX on to a single *microchip.* A digital PBX exhibits all the attributes of a computer. It is prorammable and, therefore *software controlled* has few moving parts, and is highly reliable, it is expandable, can perform extensive test and diagnostic routines, and is intelligent.

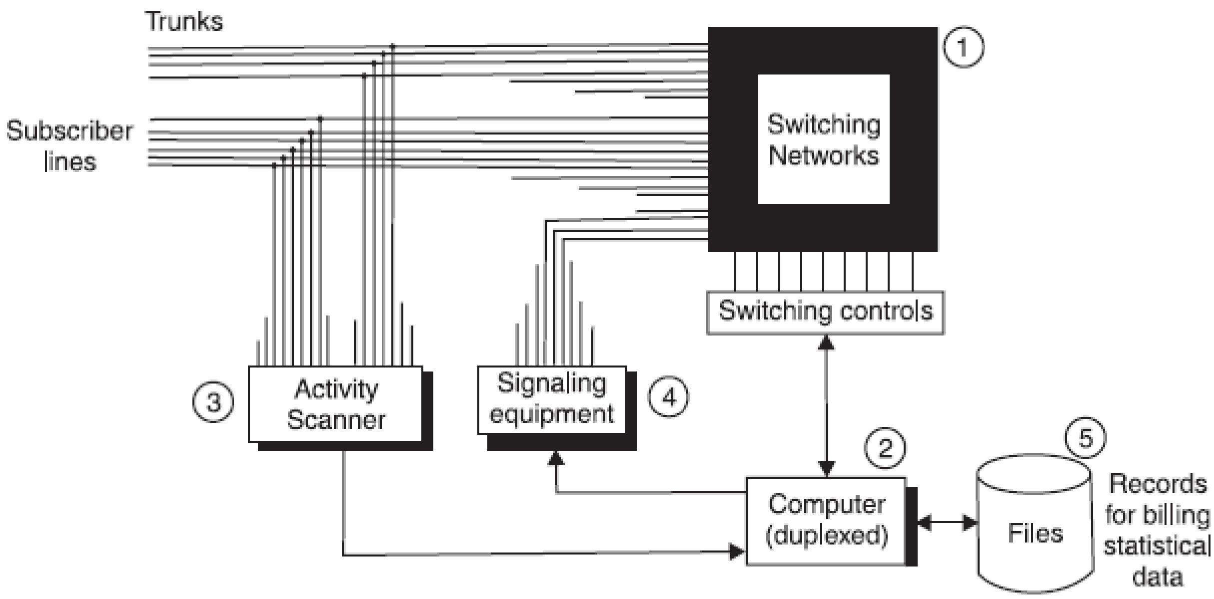


Fig. 38.17 Basic elements of a computerised switching system

FUNCTIONS OF A SWITCHING OFFICE

The *basic functions* performed by a switching office are the same, whether it is manual, electromechanical, or electronic. The *basic stages* that a call must go through are as follows :

1. When the subscriber picks up her telephone, the office must detect that service is needed. In an automatic office, the dialing tone is switched to that line, and the mechanism waits for the subscriber to dial.
2. The dialed telephone number must now be used to set up an interconnection path. The number is received as a train of pulses from a rotary dial or a train of frequency pairs from a push-button telephone. These signals cause the equipment to set up a path through the exchange to the appropriate outgoing line.
3. A central office with 10,000 lines would need approximately 50 million interconnections if there were to be a unique path between every input and every possible output. The switches therefore concentrate the calls into a limited number of paths through the exchange. Also, the office will be connected to a limited number of trunks, and again the incoming calls are concentrated onto these outgoing lines. The calls, having been concentrated onto the available interconnections, find their way to the requisite part of the exchange and are switched to the appropriate output lines.
4. The required output line might be busy. It is, therefore, necessary to detect a busy condition and to notify the caller of it. Similarly, as there is only a limited number of paths through the exchange, the exchange itself may not be capable of making the connection and again the caller must be notified. If the exchange is unable to make a connection, it will switch a busy signal onto the caller's line.
5. It is necessary to make the telephone of the called person to ring. The terminating automatic exchange sends a ringing signal down the line when the connection is made.
6. The telephone of the called party has now rung. The ringing signal must be removed from the line when that person answers. The exchange may, after a respectable wait, disconnect the call.
7. When the call is successfully established and completed, the parties put their telephones down and the circuit must then be disconnected. The exchange circuitry detects that the telephones are back on their rests. The exchange disconnects the circuit freeing the interconnection paths.
8. Last, the caller usually has to be charged. In an exchange there must be some mechanical way of recording the number of calls each subscriber makes and the duration and distance of trunk calls. Some exchanges have counters for each subscriber, which must be read periodically. Computerised exchanges produce magnetic tape for processing on another computer.

HANDS FREE PHONES VS SPEAKERPHONES

Most electronic telephones have a feature called *hands free*. These telephones have a small speaker that serves as a receiver if the phone is *on hook.* After pushing a button to access a line, dial tone is heard in the speaker. If the phone is taken *off hook*, the receiver of the handset will replace the speaker. Thus you can listen to the progress of a call being established through either the handset or the hands free speaker. The speaker will allow you to hear the called phone ringing. You will also hear the voice of the called party when he or she answers, but that person cannot hear you through the speaker. You must speak into the transmitter of the handset. *A hands free telephone allows you to place a call without picking up the handset, but once the call is connected, you must talk into the transmitter of the handset.*

Some people erroneously think that the speaker on a hands free telephone will also transmit your voice, but it will not. *It is hands free in a receive mode only*. The telephone that will transmit and receive in hands free mode is called a *speaker phone*. The speaker phone has a speaker, that acts as a receiver and also has a small transmitting device usually placed on the front edge of the telephone. You can recognise the location of this transmitter. It will be covered by the case of the telephone. The case has two to three small slits in the plastic housing to allow voice signals to reach the transmitter. *The hands free telephone allows you to listen without lifting the handset, whereas the speaker phone allows you to both talk and listen without lifting it.* Some speaker phones have electronic circuitry to detect when the user is talking. If the person is not talking, the transmitter output is very low. When the transmitter detects a voice, it reduces the output of the speaker. The operation of this electronic circuit prevents feedback to the speaker and eliminates squealing from it.

FEATURE PHONES AND ANSWERING MACHINES

When *additional features* have been added to the basic single-line telephone, such as last number redial, the phone is called a *feature phone.* The feature phone has a microprocessor and memory contained in the telephone but is still powered by the standard 52 V dc voltage from the central exchange. This electronic phone can be designed to include special features such as an *answering machine.* The answering machine uses an electronic solid state *ring detector*, which signals the microprocessor on each ring received. The microprocessor will count the number of rings received and can be programmed to answer on two or four rings, or not answer at all. The answering machine also includes a DTMF (Dual Tone Multi Frequency) receiver chip. The owner can call ino the answering machine from a remote location and use the DTMF dial, at the remote location, to provide directions to the answering machine. The directions are usually given by sending two digits as a *code.* The DTMF receiver in the answering machine receives the tones sent by the DTMF dial and is programmed to provide a specific function depending on the code received.

Fig. 38.18 30 micro cassette answering device

When the answering machine answers a call it is programmed to look at the first code as a *password.* If it receives the correct password it will honour any additional functions requested. These functions can include any one of the following; turn on the machine, turn off the machine, playback all messages, playback only new messages, record a new announcement, fast forward, fast reverse, rewind, change the password, and so on. *The use of DTMF tones to signal the answering machine allows a person to perform any administrative function on the machine from any remote location.* The machine has programming that will automatically answer an incoming call, wait for and recognise a password code, return a record announcement, respond to DTMF signals and intepret the received code and perform the function assigned to that code.

EXERCISES

Descriptive Questions

1. What is the need for a switching system?
2. With the help of a suitable sketch, briefly explain the working of a selecting switch.
3. Explain the significance of central points and exchange.
4. Write short notes on :
    a. Manual exchange
    b. Automatic exchange
    c. Computer-controlled exchange
    d. Call-accounting equipment
    e. Step-by-step principle
    f. Uniselector
5. Explain the difference between a rotary-dial telephone and a push-button telephone.
6. With the help of a suitable sketch, explain the principle of switching by electromechanical uniselector of one-from-ten outlets.
7. Differentiate between a uniselector and a two-motion selector.
8. Explain a three-digit numbering scheme with the help of a two-motion selector.
9. What are the features of step-by-step selectors?
10. List the drawbacks of step-by-step electromechanical exchanges.
11. Explain the principle of operation of a reed relay.
12. What are the advantages of crossbar system?
13. With the help of a suitable sketch, explain, the principle of switching by a 4 × 4 matrix switch.
14. Define and explain stored program control (SPC).
15. Explain PBX switching.
16. With the help of a suitable sketch explain the basic elements of a computerised switching system.
17. What are the functions of a switching office?
18. Differentiate between a hands free phone and a speaker phone.
19. How does an answering machine operate?
20. What role is played by DTMF tones in answering machines?

Fill in the Blanks

1. A switching system enables any terminal to pass information to any other terminal as ___________________ by the calling customer.
2. So far as telephone networks are concerned, it is more economical to perform all switching functions at ___________________.
3. The whole of the telephone network has to be designed to strict ___________________ parameters so that any subscriber on any exchange can converse satisfactorily with all other subscribers anywhere in the world.
4. The wiper rotates in one plane only, so this type of electromechanical switch is called a ___________________
5. In a two-motion selector, the wiper first moves ___________________ to the appropriate level, and then rotates ___________________ to a particular contact on that level.
6. Each digit dialed takes the subscriber one step ___________________ to the called number.
7. A large amount of expensive equipment in step-by-step exchanges is ___________________ most of the time.
8. Elecromechanical step-by-step exchanges have a high cost of ___________________.
9. Exchanges using a register marker are called ___________________ exchanges.
10. A crossbar exchange is more ___________________ than a step-by-step exchange.
11. Very little ___________________ is needed in crossbar exchanges.
12. Fewer ___________________ men are required in crossbar exchanges.
13. The intersections between horizonal and vertical lines are called ___________________.
14. Efficiency can be improved by using smaller matrix switches ___________________ together.
15. The use of computers to conrol telephone exchange ___________________ is called stored program control.
16. The PBX is essentially a ___________________ device.
17. A digital PBX can handle both voice and ___________________ traffic.
18. The speaker phone allows you to both ___________________ and ___________________.

ANSWERS

Fill in the Blanks

1.	selected
2.	central points
3.	transmission
4.	uniselector
5.	vertically, horizontally
6.	nearer
7.	idle
8.	maintenance
9.	common control
10.	reliable

11. maintenance
12. exchange
13. crosspoints
14. linked
15. switching
16. connecting
17. data
18. talk, listen

CHAPTER 39

MODULATION TECHNIQUES

Over the years modulation methods have been devised for transmitting the required information as effectively as possible with the minimum amount of distortion. The primary factors to be considered are signal power, bandwidth, distortion and noise power. Ultimately, it is the ratio of signal power to noise power or output signal to noise ratio specified for the system which determines its performance. Consequently, it is not surprising to find a wide range of modulating techniques being used which appear to compete with one another under given practical conditions. Broadly speaking, these various techniques may be grouped into analog methods which use a sine wave as the carrier signal and pulse method which use a digital or pulse train as the carrier signals.

ANALOG METHODS

The two most important analog methods are amplitude modulation and angle modulation. *Amplitude modulation* (AM) with both sidebands and carrier present is most common for certain application, such as radio broadcasting and radio telephony (see Fig. 39.1).

More *economical versions* of AM are vestigial sideband transmission (VSB) which is used in television for economising bandwidth, while double sideband suppressed carrier (DSBSC) or single sideband suppressed carrier (SSBSC) provide further power or bandwidth economy. In particular, SSBSC is used extensively in multiplex coaxial systems for carrying several messages simultaneously. *However the AM systems are essentially narrowband systems and suffer from limitations due to noise which has a direct effect on signal amplitude.*

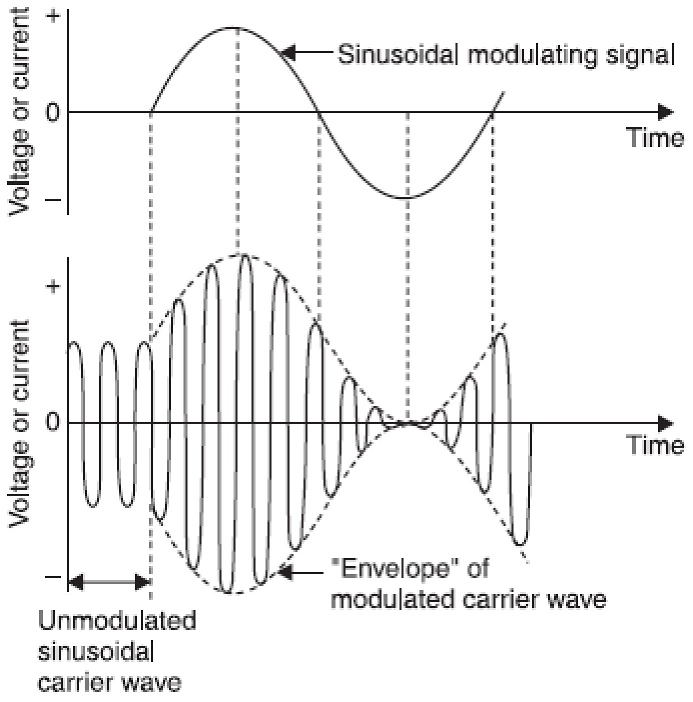


Fig. 39.1 Amplitude modulated carrier wave

Some systems use *angle modulation* because of its immunity to amplitude-varying noise. In angle modulation, the instantaneous angle of the carrier wave is varied and this leads to two forms of modulation known respectively as *frequency modulation* (FM) and *phase modulation* (PM). As a consequence, FM and PM are closely related though practical systems tend to favour FM rather than PM Typical examples are VHF communication, satellite communication and FM radar. However, because the FM carrier wave, shown in Fig. 39.2, requires a much greater bandwidth than its AM counterpart, *an FM system is capable of giving a much improved signal to noise performance compared to that of the corresponding AM system* or, alternatively, a considerable economy on power, if required. Hence, FM systems are to some extent superseeding AM systems.

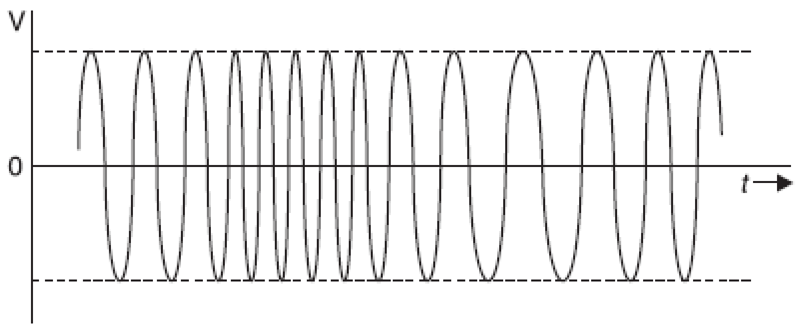


Fig. 39.2 Frequency modulated carrier wave

Amplitude modulation : A bandwidth of 300 to 3400 Hz is required for the transmission of commercial quality speech. To economise on cable it is desirable to be able to transmit more than one conversation over a single pair of wires. If several conversation signals were all connected together at one end of a line, it would not be possible to *separate* them at the distant end since each conversation would be occupying the *same frequency spectrum* of 300 Hz to 3400 Hz. Amplitude modulation (AM) plus frequency division multiplexing (FDM) is one way of solving this problem. Each conversation is *shifted* to a different part of the frequency spectrum by using a high-frequency waveform to *carry* each individual speech signal. These high-frequencies are called *carrier frequencies.*

It can be shown that, when a sinusoidal carrier wave of frequency f_c Hz is amplitude modulated by a sinusoidal modulating signal of frequency f_m Hz, then the *modulated carrier wave* contains three frequencies: the *original* carrier *frequency*, f_c Hz, the *sum* of the carrier and modulating frequencies $(f_c + f_m)$ Hz and the *difference* between carrier and modulating frequencies, $(f_c - f_m)$ Hz. This is illustrated in Fig. 39.3.

Two of these frequencies are new, being produced by the amplitude modulation process, and are called *side frequencies :* the *upper* side frequency (sum) and the lower side frequency (difference). The *bandwidth* of the modulated carrier wave is $(f_c + f_m) - (f_c - f_m) = 2f_m$, i.e. double the modulating signal frequency.

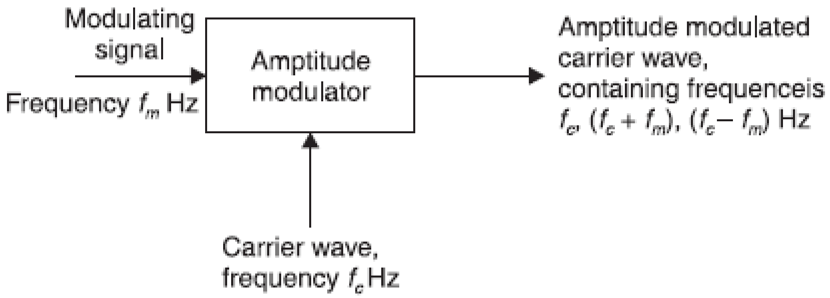


Fig. 39.3 Principle of amplitude modulation

When the modulating signal consists of a band of frequencies, then each individual frequency will produce upper and lower side-frequencies about the unmodulated carrier frequency and so upper and lower *sidebands* are obtained. This is shown in Fig. 39.4.

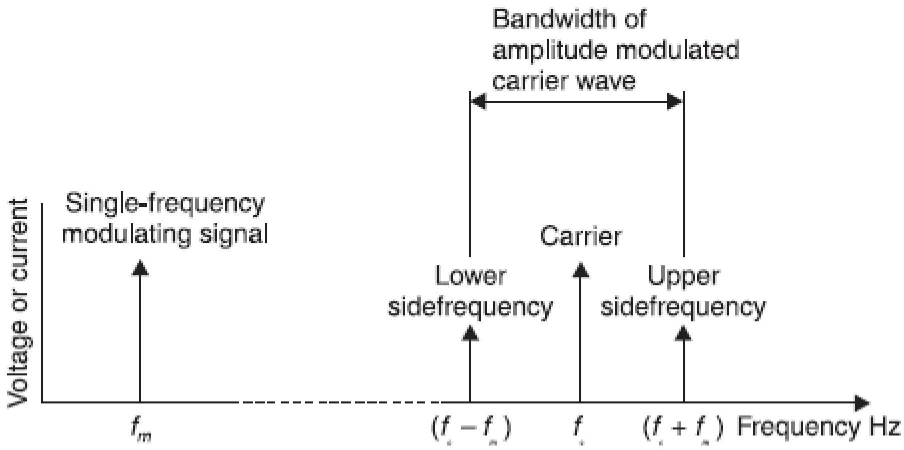


Fig. 39.4 Frequency spectrum of an amplitude-modulated wave for single-frequency modulating signal

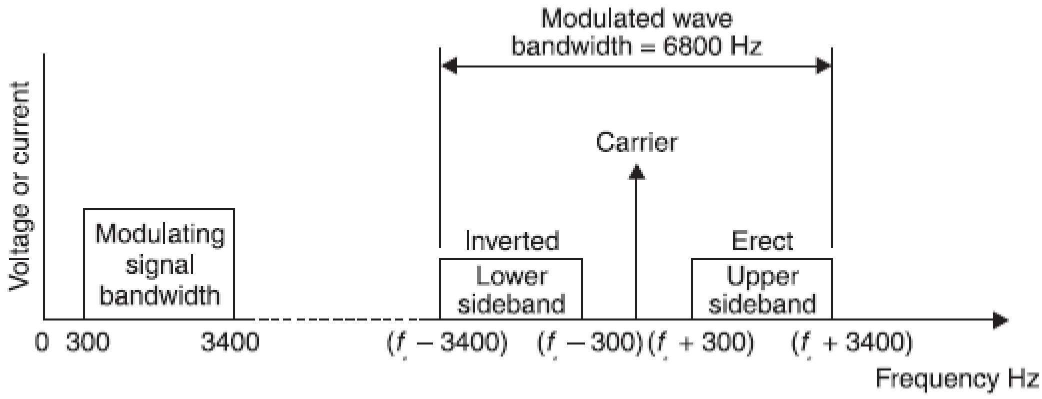


Fig. 39.5 Frequency spectrum an amplitude-modulated wave for commercial speech modulating signal

As the modulating signal bandwidth increases, the modulated wave bandwidth also increases, and the transmission system must be *capable of handling* this bandwidth throughout. All the information is carried by *either one* of the sidebands. The carrier component is of constant amplitude and frequency and hence does not carry any of the information signal at all.

It is possible by using special equipment to *suppress* both the carrier and one sideband and to transmit just the other sideband with no loss of information. This is called *single sideband working* (SSB) or single sideband *suppressed carrier* working. This method is *complex and costly* and hence it is not used for domestic radio broadcasting but it is used for some long-distance radio telephony systems and for multi-channel carrier systems used in national telephone networks.

Frequency Modulation : Another method of superimposing information signals on to a carrier is frequency modulation in which the modulating signal varies the frequency of a carrier wave.

The concept of FM can best be understood by considering a modulating signal of *rectangular waveform* such as the one shown in Fig. 39.6. Suppose the unmodulated carrier frequency is 3 MHz. The periodic time of the carrier voltage is 1/3 μs and so three complete cycles of the *unmodulated carrier wave* will occur in 1 μs. When after 1μs, the voltage of the modulating signal increases to +1 V, the instantaneous carrier frequency *increases* to 4 MHz. Hence in the time interval 1 μs to 2 μs, there are four complete cycles of the carrier voltage. After 2 μs the modulating signal returns to 0 V and the instantaneous carrier frequency falls to its original value, 3 MHz. During the time internal 3 μs to 4 μs, the modulating signal voltage is –1 V and the carrier frequency *reduces* to 2 MHz; this means that two cycles of the carrier voltage occur in this period of time. When, after 4 μs, the modulating voltage is again 0 V, the instantaneous carrier frequency is *restored* to 3 MHz. At t = 5 μs, the modulating voltage is +2 V and, since *frequency deviation is proportional to signal amplitude*, the carrier frequency is deviated by 2 MHz to a new value of 5 MHz. Similarly, when the modulating signal voltage is – 2 V, the deviated carrier frequency is 1 MHz. *At all times the amplitude of the frequency modulated carrier wave is constant at* 1V, *and this means that the modulation process does not increase the power content of the carrier wave.*

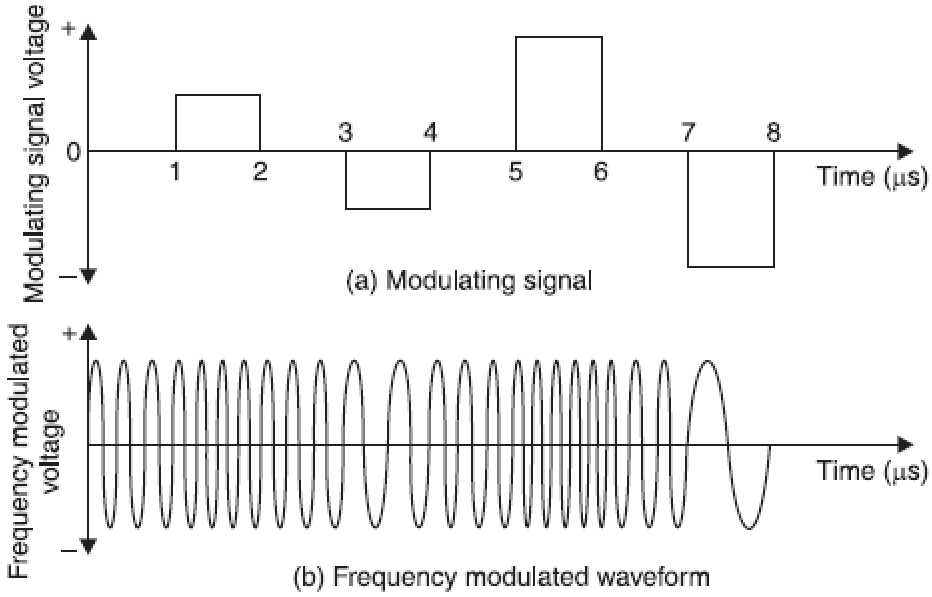


Fig. 39.6 Frequency modulation of a rectangular wave

When the modulating signal is of sinusoidal waveform, the frequency of the modulated carrier wave will also vary sinusoidally as shown in Fig. 39.7. *There is no inherent maximum value of the frequency deviation that can be obtained in a frequency-modulation system*, this should be compared with amplitude modulation where the maximum amplitude deviation possible corresponds to 100% modulation i.e. reduction of the amplitude of the envelope to zero (see Fig. 39.1).

FM has a number of *advantages* over AM. The price which must be paid for some of the advantages of FM over double sideband AM is a *wider bandwidth* requirement. FM is used for sound broadcasting in the VHF band, for the sound signal of television broadcasting, for some mobile systems, and for multi–channel telephony.

A form of modulation which uses a digital data signal to modulate a sine-wave carrier is called digital modulation. The three types mainly used are *amplitude-shift keying* (ASK), *frequency-shift keying* (FSK) and *phase-shift-keying* (PSK). They correspond approximately to AM, FM, and PM and are used especially in data communication systems.

ASK : *On/off keying* is the earliest modulation method. A continuous radio frequency wave (CW) is *interrupted in a recognizable pattern* (Morse code). To provide audibility the carrier is heterodyned with a beat frequency oscillator (BFO) in the receiver. The use of a modulated continuous wave (MCW) *eliminates* the need of a BFO but the bandwidth of the signal is increased. *The problem with on/off keying is the lack of a reference level.* If the signal strength temporarily falls below the sensitivity threshold of the receiver it appears to the operator as a series of spaces.

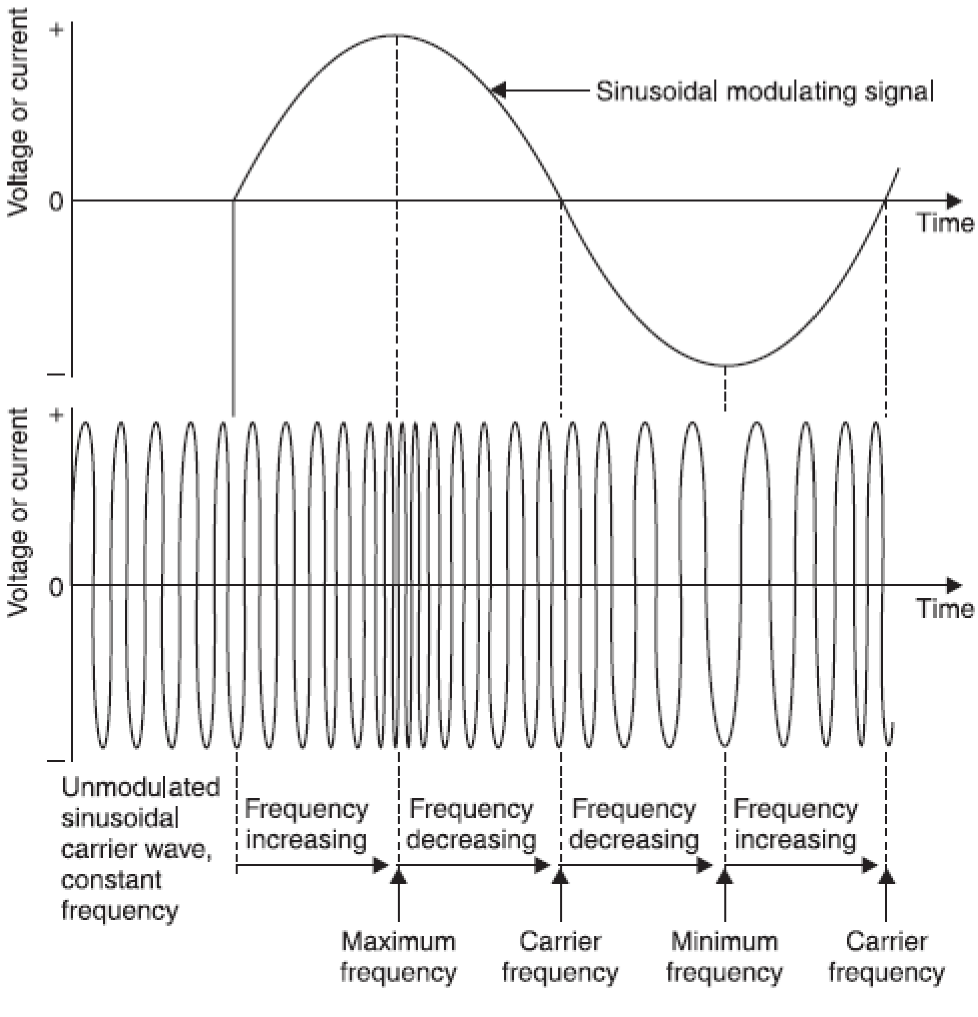


Fig. 39.7 Frequency modulation of a sinusoidal waveform

Binary Amplitude Shift Keying (BASK) : This shifts the level of an audio frequency *subcarrier* which then modulates a radio frequency *carrier*. Because the level of a subcarrier is changed, *AM sidebands* are produced. Also, because the keyed waveform is non-sinusoidal *harmonics* occur. The occupied sub-carrier bandwidth for ASK is:

$$\text{Bandwidth} = 2B \qquad \ldots(39.1)$$

where B is the bit repetition rate (bits/second).

When the RF carrier is modulated its bandwidth is 2 (f_c + B) where f_c is the subcarrier frequency.

FSK : Although used for conveying *digital information*, frequency shift keying in reality employs frequency modulation. In its original form, developed for HF frequency transmission, FSK changes the carrier frequency to indicate a 1 or 0 but retains the nominal carrier frequency as a reference and to represent a *mark*. A downward shift of carrier frequency by 170 Hz represents a *space* in the HF radio system.

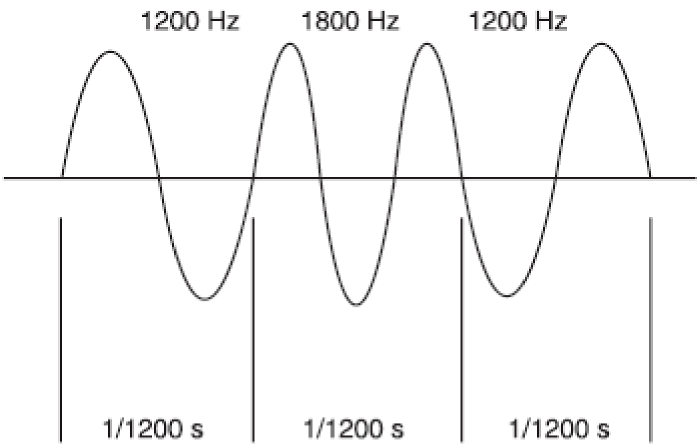


Fig. 39.8 Fast frequency shift keying modulation

Modern FSK uses two different modulation frequencies to represent 1s and 0s. If the *intersymbol interference* (ISI) is to be avoided the separation of the tones must be more than half the bit rate, and a factor of 0.7 is often used.

The base bandwidth requirement is :

$$\text{Bandwidth} = f_2 - f_1 + 2B \qquad \ldots(39.2)$$

The bandwidth of a modulated carrier is :

$$\text{RF bandwidth} = 2(f_2 + B) \qquad \ldots (39.3) \text{ (narrow band FM)}$$

Minimum shift keying is a form of FSK where the frequency deviation is equal to half the bit rate.

Fast Frequency Shift Keying (FFSK) may either amplitude or frequency modulate the carrier. In binary FFSK, the data is changed in a modem to tones of 1800 Hz to represent binary 0 and 1200 Hz to represent binary 1. During transmission a binary 1 consists of 1 cycle of 1200 Hz, and a 0 consists of 1½ cycles of 1800 Hz (f_1 and f_2) i.e. a bit rate of 1200 bps. For acceptable intersymbol interference the distance between the tones cannot be less than half the bit rate and the 600 Hz separation in FFSK represents the *fastest signalling speed*—hence the description—and *minimum bandwidth*. For this reason it is sometimes called minimum frequency shift keying (MFSK). The base bandwidth is the same as for FSK, but the RF bandwidth depends upon the system deviation.

PSK : There are several variants of phase shift keying. Binary phase shift keying changes the phase of the carrier by 180° at the zero crossing point (Fig. 39.9). No carrier frequency is present with PSK as half the time the carrier is multiplied by + 1 and the other half by –1 and cancels out, but *the reference phase of the carrier must be re-inserted at the receiver*. The bandwidths occupied are the same as for ASK i.e.

$$\text{Baseband} = 2B$$
$$\text{RF bandwidth} = 2(fc + B)$$

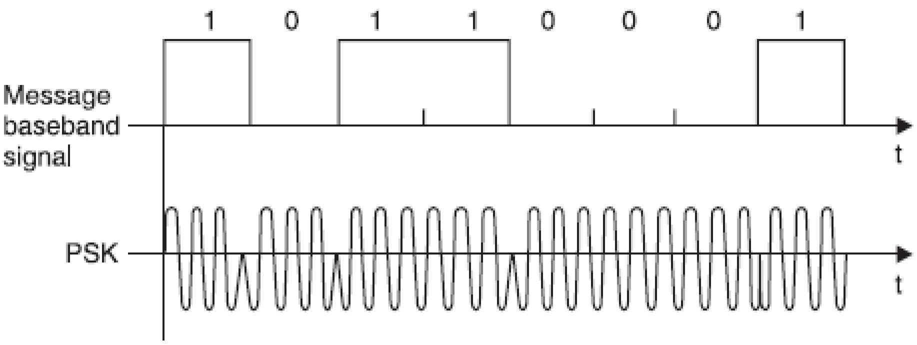


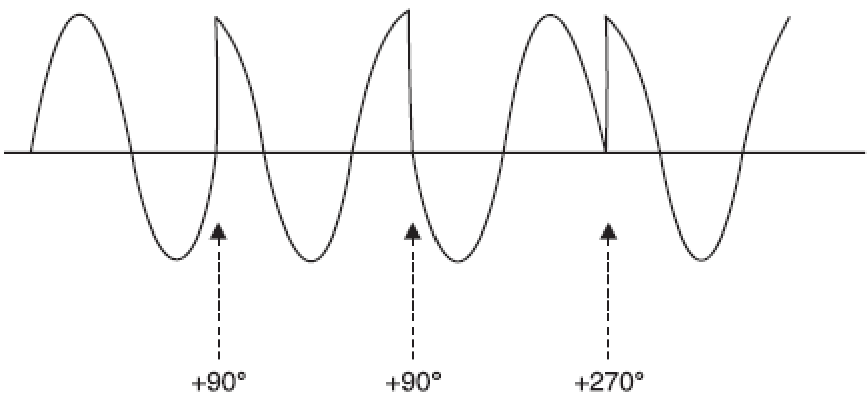
Fig. 39.9 Phase shift keying modulation

Differential phase shift keying (DPSK) advances the phase by 90° or 270° at each change of logic state (Fig. 39.10). *Changing phase only at a change of logic state saves bandwidth* which, for DPSK, is equal to the bit rate.

An important advantage of both FFSK and PSK over FSK is that *because the moment of change is predefined, it is possible to recover data more accurately. However, the transition between signalling states is not smooth requiring large and rapid phase shifts.* Multilevel systems with less phase shift between elements are preferable.

PULSE METHODS

An alternative method of modulation uses a digital carrier signal comprising a *pulse train* which can be modulated to carry the required information. The amplitude, width or position of the pulses can be altered by the information signal as shown in Fig. 39.11. *Pulse amplitude modulation* (PAM) is that form of modulation in which the amplitude of the pulse carrier is varied in accordance with some characteristic, normally the amplitude, of the modulating signal. *Pulse width modulation* (PWM) is that form of modulation in which the duration of a pulse is varied in accordance with some characteristic of the modulating signal. *Pulse position modulation* (PPM) is that form of modulation in which the positions in time of the pulses are varied in accordance with some characteristic of the modulating signal without a modification of pulse width.



Data	Phase change
0	+90°
1	+270°

Fig. 39.10 Differential phase shift keying modulation

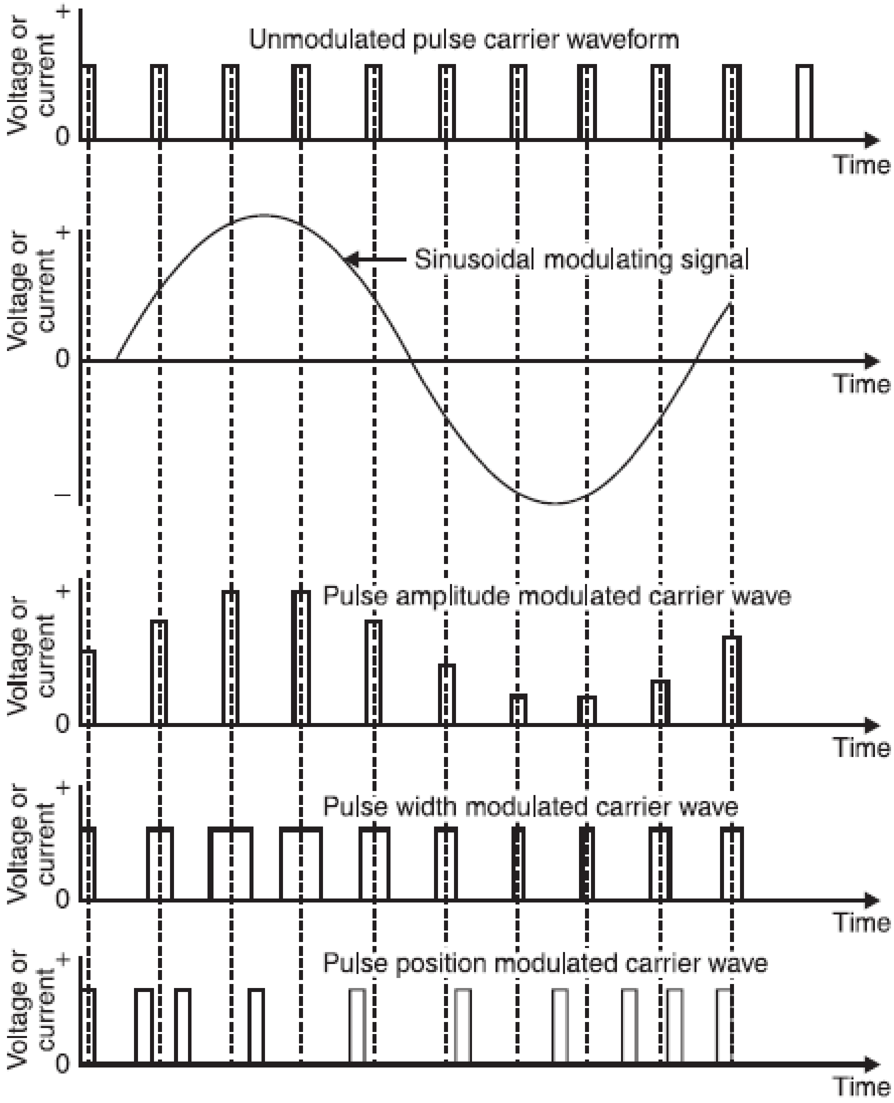


Fig. 39.11 Pulse-modulated carrier waves

Pulse Code Modulation (PCM) : In a PCM system the analog signal is *sampled at regular intervals* to produce a pulse amplitude modulated waveform. If an analog signal is sampled regularly using a *sampling rate* of atleast twice the highest frequency of the signal, the samples are found to be adequate to allow the recreation of the original voice signal with *sufficient accuracy* for all practical purposes. *Sampling* is done by feeding the analog signal to a circuit with a gate which opens only for the duration of the sampling pulse. The output is a *pulse amplitude modulated (PAM) signal.* This is shown in Fig. 39.12.

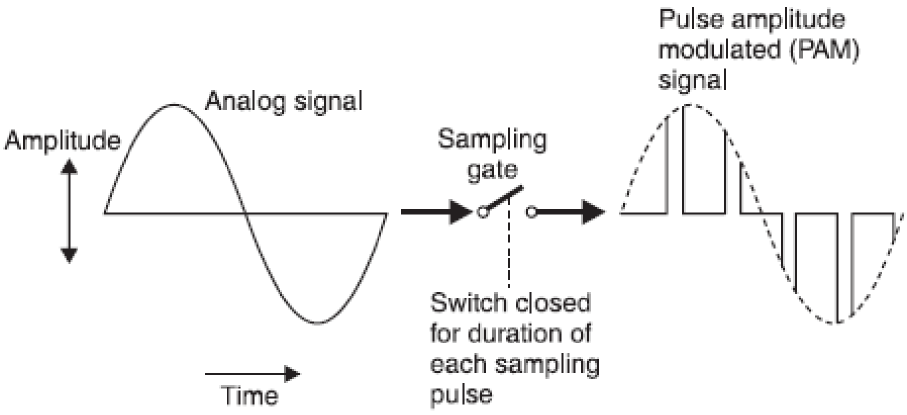


Fig. 39.12 Principle of pulse amplitude modulation

Figures 39.13 and 39.14 show the effect of taking these samples from sound signals of different voice frequencies. Although the commercial voice band goes upto 3400 Hz, almost all the power of human speech is at much lower frequencies (for example at around 500 Hz for male voices) so *several samples are taken during each cycle, enough to enable the original analog signal to be reconstituted with a fair degree of accuracy.*

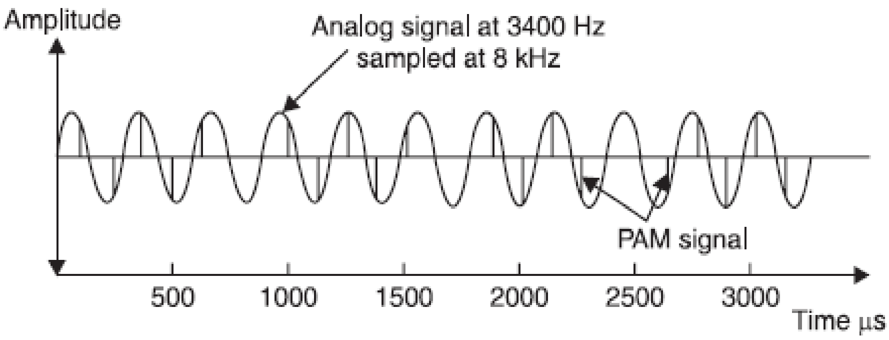


Fig. 39.13 Sampling of the highest voice frequency to be transmitted

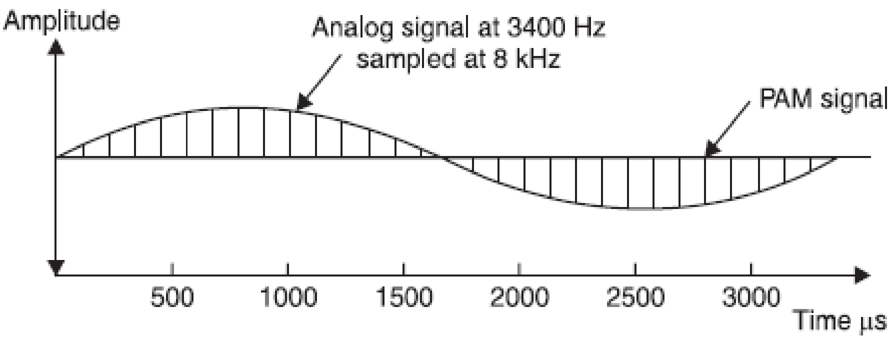


Fig. 39.14 Sampling of the lowest voice frequency to be transmitted

The total amplitude range that the signal may occupy is divided into a number of *levels*, each of which is allocated a number. There are 256 different levels in internationally recommended PCM.

There are two different *PCM encoding* laws in use in the world, called μ-law coding (developed and used in America) and A-law coding (developed and used in Europe). These two laws give different quantization values for a signal of a given amplitude so that *PCM channels to these two standards cannot directly interwork unless special interfacing equipment is provided.*

The variable-amplitude PAM signal is then *compared* with the instantaneous value of the appropriate range of levels which are called *quantizing intervals.* The signal is assigned the value of the interval in which it falls and the number of this value is then *encoded* into an 8 digit binary code. This gives 2^8 or 256 possible levels, 128 each side of zero. Each 8 digit code describing the amplitude level of the sample is known as a *PCM word.* Since the sampling rate is 8000 times per second and each sample has 8 digits, there are 64000 bits per second for a single PCM channel.

Fig. 39.15 shows how this *quantization process* works for a complex analog signal. Only 8 sampling levels are shown for clarity; these 8 levels require only 3 binary digits against the 8 binary digits needed to distinguish between the 256 levels used is practical PCM system.

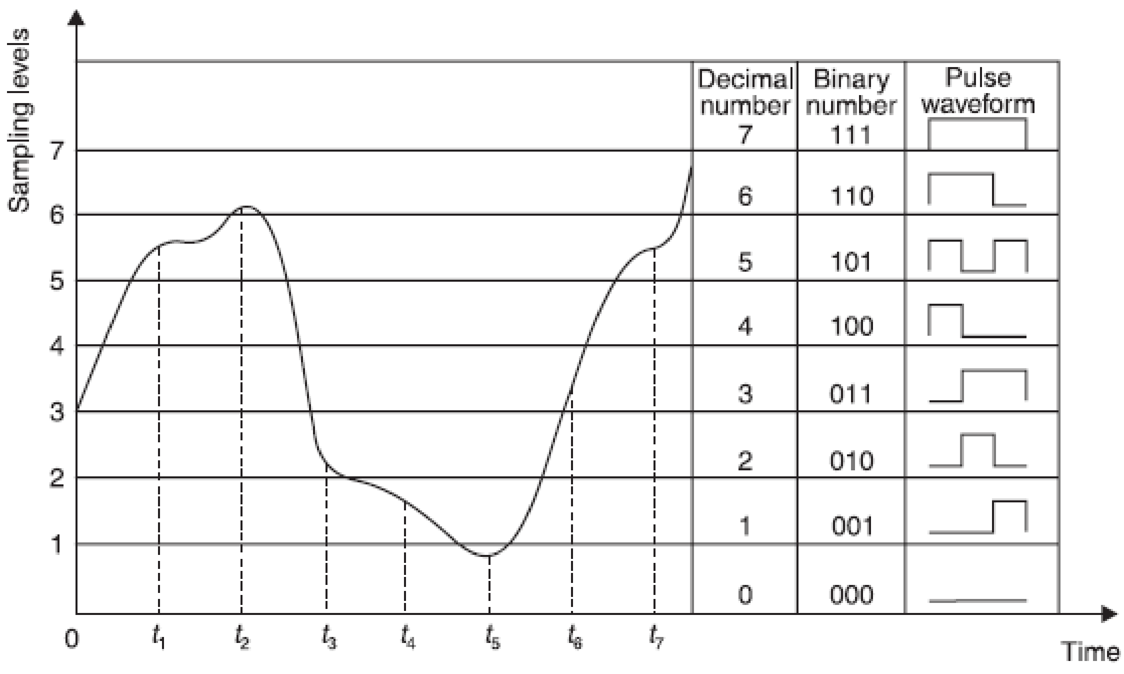


Fig. 39.15 Quantization of a signal

The signal waveform is *sampled* at time instants t_1, t_2, t_3 etc. At time t_1 the instantaneous signal amplitude is in between levels 5 and 6 but since it is nearer to level 6, it is approximated to this level. At instant t_2 the signal voltage is slightly greater than level 6 but is again rounded off to that value. Similarly the sample taken at t_3 is represented by level 2, the t_4 sample by level 2, the t_5 sample by level 1, and so on. The binary pulse train which would be transmitted to represent this signal is shown in Fig. 39.16. A space, equal in duration to one binary pulse, has been left in between each binary number in which *synchronisation information* can be transmitted.

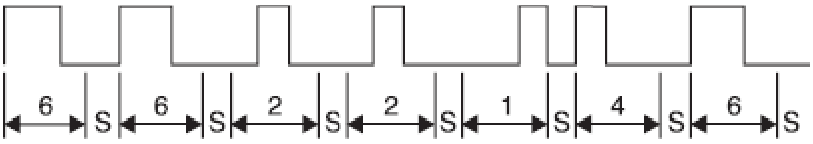


Fig. 39.16 Binary pulse train representing the signal shown in Fig. 39.15

A PCM system transmits signal information in digital form. The quantization process will result in some error at the receiving end of the system when the analog signal is reconstituted. The error appears in the form of *quantization noise* and can be reduced only by increasing the number of sampling levels. This would increase the number of binary digits required and the *bandwidth* which must be provided would be wider. Practical systems have to accept a compromise solution. *With an 8 binary digit scheme, the quantization noise is generally considered acceptable.*

For the receiving equipment to be able to decode the incoming binary pulse trains, it is only necessary for it to be able to determine whether or not a pulse is present. The processes of encoding and quantizing are *reversed:* a replica of the PAM signal is first generated, then this is fed to a low pass filter to reconstruct the original analog signal.

MULTIPLEXING

All transmission media have capacities great enough to carry more than one voice channel. In other words, their bandwidths are considerably greater than the 3 kHz needed for transmitting the human voice. At the *top end* of the scale microwave and fibre-optic circuits carry thousands of voice channels; at the *lower end* of the scale, each voice channel may be split into 12 or 24 telegraph channels.

Where a facility is set up, such as a chain of microwave links, which has a *broad bandwidth* it is very desirable to make the maximum use of this bandwidth by making it carry as many channels as possible. It is often desirable to construct a communication link with as wide a bandwidth as possible and then *divide* the bandwidth between as many users as possible. Many separate signals are *multiplexed* together so that they can travel as one signal over a high bandwidth.

In a *multiplex system*, two or more signals are *combined* so that they can be *transmitted together* over one physical cable or radio link. The resulting combined signal is transmitted over a system with a suitably high bandwidth. When it is *received*, it must be *split up* into the separate signals of which it is composed.

The word *multiplexing*, in general, means the use of one facility to handle *several separate but similar* operations simultaneously. In telecommunication language it means the use of one telecommunication link to handle several channels of voice or data. Multiplexing is possible because the operations that are multiplexed take place at a considerably slower speed than the operating speed of the facility in question. *Multiplexing is a key factor in effectively utilising existing telecommunication links.*

Communication channels are normally grouped together in *packages* that fill the bandwidth available on different types of plant. High capacity links can be created by multiplexing large numbers of channels that were, themselves, grouped together stage by stage. The original signal may go through many *multiplexing* stages and equivalent *demultiplexing.* It is thus worked upon by a variety of *electronic conversion processes* before it ultimately arrives at its destination.

There are three methods of transmitting more than one signal over one path: space division multiplexing, frequency division multiplexing and time division multiplexing. *Space division multiplexing* means that more than one physical transmission path are grouped together. Wire-pair cables, for example, are constructed containing many hundreds of wire pairs. Coaxial cables contain 20 or so tubes, giving a high total bandwidth. *Frequency division multiplexing* and *time division multiplexing* are alternate techniques for splitting up a single physical path. The quantity of information that can be carried is proportional to the *range of frequencies* (the bandwidth) used and to the *period of time* used. If the quantity of information required from one channel is less than that which the physical facilities can carry, then the space available can be divided up either in *frequency slices,* or *time slices.*

In either case, the engineering limitations of the devices employed prevent the slices from being packed tightly together. With frequency division a *guard band* is needed between the frequencies used for separate channels, and with time division, a *guard time* is needed to separate the time slices. If the guard bands or guard times were made too small, the expense of the equipment would increase out of proportion to the advantage gained.

CONCENTRATORS AND MULTIPLE ACCESS

When a fixed number of facilities are available for use they can be assigned to users in a fixed or variable manner. In today's telephone systems, the local loops are normally assigned on a *fixed basis*, and the trunks are assigned on a *variable basis.* The trunks are therefore used much more efficiently than the local loops.

Suppose that there are 100 subscribers in a locality who use their telephones occasionally. There is, in theory, no need for 100 channels to connect them to their local switching office. Twenty channels, for example, could be used with some means of allocating a channel to a subscriber when he needs it. This technique is called *concentration.* There are various ways in which it can be done; hence there is a variety of devices called concentrators.

There is a fundamental difference between concentration and multiplexing. With *multiplexing* all subscribers can have a channel simultaneously if they want one. With *concentration* they cannot. In the above example if all 20 channels are in use and a twentyfirst subscriber requests a channel, he will be unlucky. He either receives a busy signal or must wait until a channel becomes free. Such is the nature of concentration. *It takes advantage of the fact that not all the users are active all the time.*

The design of a concentrator depends on the *type* of signal it is to concentrate. A concentrator for telephone lines may be an electromechanical device that scans a bundle of lines searching for a free one. It may be a solid state circuit that concentrates PCM traffic to travel over a digital trunk.

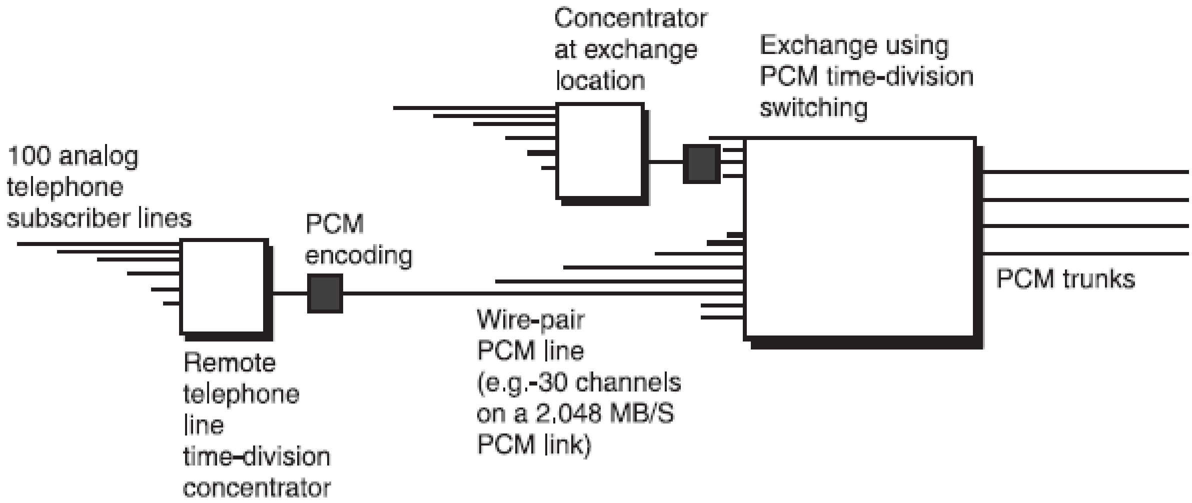


Fig. 39.17 A remote telephone concentrator designed for a PCM line

When a large number of locations are to be interconnected, switching offices are used to lower the number of *links* that are required. When high capacity channels are available at reasonable cost, there is an *alternative* to switching offices. Many geographically dispersed access points could share the same channel, with a *control mechanism* to enable them to intercommunicate as they need.

High capacity channels—such as the PCM links, digital radio, CATV cables, optical fibres and satellites are available. If they are analog channels, they can be shared by *frequency division.* If they are used in a digital fashion, they can be shared by *time-division.* Most telephone trunks are shared by point-to-point multiplexing. If there are many access points rather than two, the technique is referred to as *multiple access.* If it is done in a digital time-division fashion, it is called *time-division multiple access*, generally referred to as TDMA.

Frequency-division multiple access, FDMA, is also used. Frequency division implies that *separate frequencies* are allocated to different users, as in radio broadcasting; a transmitter or receiver must be tuned to the frequency that is assigned to it. Time division implies that *different time slots* are allocated to different users, and each user must know at what times he must transmit and receive.

TWO-CHANNEL TDM SYSTEM

The principle of a simple two-channel TDM system is illustrated in Fig. 39.18. *Analog inputs are considered for clarity*. TDM is, of course, normally a means of multiplexing digital signals, such as PCM pulses, all having the same amplitude.

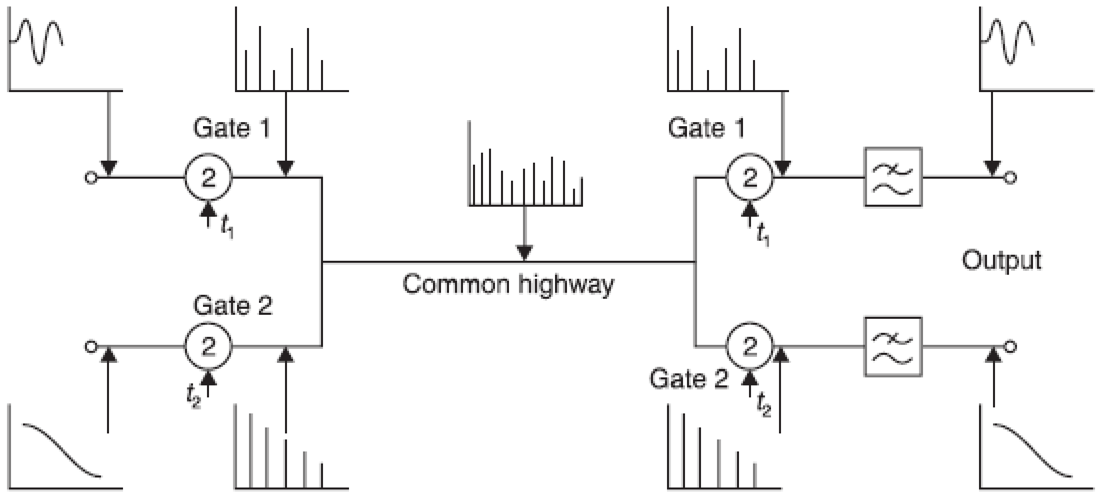


Fig. 39.18 A simple two-channel TDM system: t_1 = a series of pulses occurring at fixed intervals; t_2 = a series of pulses occurring at the same periodicity as t_1 but commencing later by an amount equal to half the time interval

The two channels, which are to share the *common circuit*, are each connected to it via a *channel gate*. The channel gates are electronic switches which permit only the signal present on a channel to pass when opened by the application of a *controlling pulse.* Hence, if the controlling pulse is applied to gate at time t_2, and not to gate 2, gate 1 will open for a time equal to the duration of the pulse but gate 2 will remain closed. During this time, therefore, a pulse or sample of the amplitude of the signal waveform on channel 1 will be transmitted to line. At the end of the pulse, both gates are closed and no signal is transmitted to line. If now the controlling pulse is applied to gate 2 at a latter time (see Fig. 39.18), gate 2 will open and a sample of the signal waveform on channel 2 will be transmitted to line. Thus if the pulses applied to control the opening and closing of gates 1 and 2 are *repeated at regular intervals*, a series of samples of the signal waveforms existing on the two channels will be transmitted.

At the receiving end of the system, gates 1 and 2 are opened, by the application of control pulses, at those instants when the incoming waveform samples, appropriate to their channel, are being received. This requirement demands *accurate synchronisation* between the controlling pulses applied to the gates 1, and also between the controlling pulses applied to the gates 2. If synchronisation signals are sent from one end to the other as an integral component of the PCM system (as they are with internationally specified systems), all the signals will, of course, maintain their *correct relative positions.* If the pulse synchronisation is correct, the waveform samples are directed to the correct channels at the receiving end. The received samples must then be covered back to the original waveform, i.e. *demodulated.*

In its passage along a telephone line, the TDM signal is both *attenuated and distorted* but, provided the receiving equipment is able to determine whether a pulse is present or absent at any particular instant, no errors are introduced. To keep the pulse waveform within the accuracy required, pulse regenerators are fitted at intervals along the length of the line. The function of a pulse regenerator (Figs. 39.19 and 39.20) is to check the incoming pulse train at accurately timed intervals for the *presence or absence* of a pulse. Each time a pulse is detected, a new undistorted pulse is transmitted to line and, each time no pulse is detected, a pulse is not sent.

In Fig. 39.19 the bit stream is first *equalised* and then *amplified* to reduce the effects of line attenuation and group-delay/frequency distortion. The amplified signal is applied to a *timing circuit* which generates the required timing pulses. These timing pulses are applied to one of the inputs of *two-input* and *gates,* the phase-split amplified signal being applied to the other input terminals of the two gates. Whenever a timing pulse and a peak (positive or negative) of the incoming signal waveform *occur at the same time,* an output pulse is produced by the appropriate pulse generator. It is arranged that an output pulse will not occur unless the peak signal voltage is greater than some *pre-determined value* in order to prevent false operation by noise peaks. *The use of pulse regenerators allows very nearly distortion free and noise free transmission regardless of the route taken by the circuit or its length.*

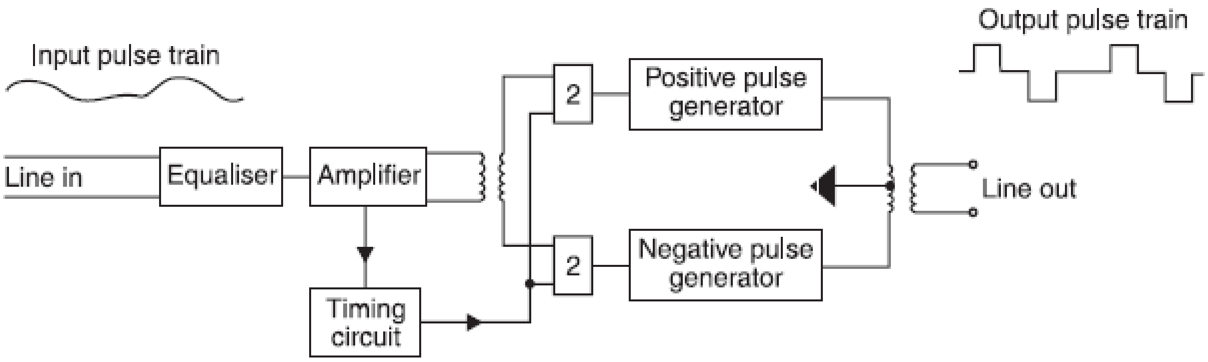


Fig. 39.19 Pulse regenerator

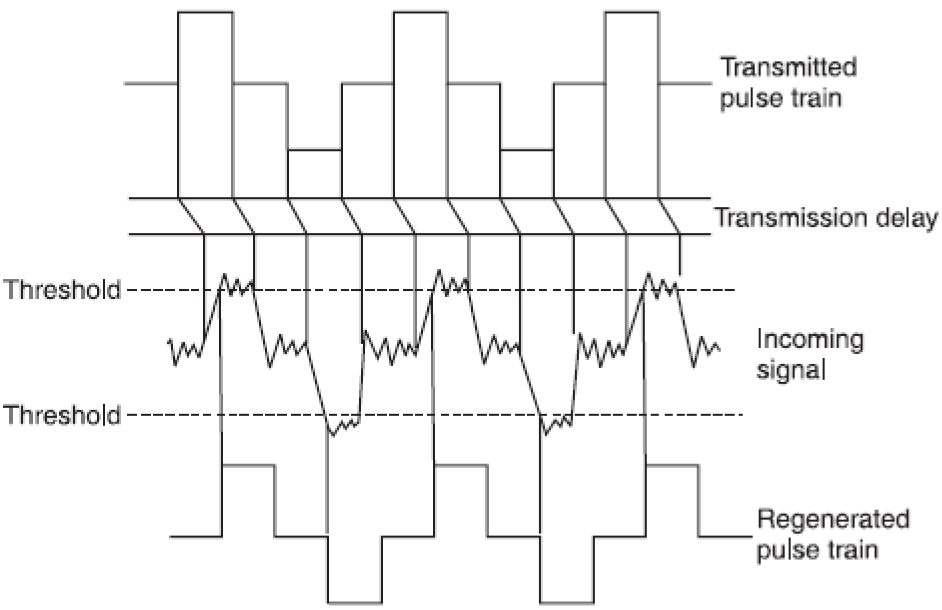


Fig. 39.20 Waveform regeneration by a pulse regenerator

EXERCISES

Descriptive Questions

1. Briefly explain the analog methods of modulation.
2. Differentiate between AM and FM.
3. Explain the relative advantages and disadvantages of FM over AM.
4. With the help of a suitable sketch explain the frequency spectrum of an amplitude modulated wave for single-frequency modulating signal.
5. What are the digital methods of modulation? Explain briefly.
6. Differentiate between ASK, FSK, and PSK.
7. What are the pulse methods of modulation? Explain briefly.
8. Describe the process of quantization of a signal.
9. Why multiplexing is required? How is it achieved?
10. Differentiate between SDM and TDM.
11. Explain the difference between concentration and multiplexing.
12. How pulse regeneration is achieved?

Fill in the Blanks

1. VSB transmission is an ________________ version of AM.
2. AM systems are ________________ band systems.
3. AM systems suffer from limitations due to ________________.
4. Noise has a ________________ effect on signal amplitude.
5. An FsM system give a much improved ________________ performance as compared to an AM system.
6. Frequency division is proportional to signal ________________.
7. The FM process does not increase the ________________ content of the carrier wave.
8. FM systems are ________________ band systems.
9. The problem with on/off keying is the lack of a ________________.
10. If inter-symbol interference is to be avoided, the separation of the tones must be more than ________________ the bit rate.
11. 600 Hz separation in FFSK represents the ________________ signalling speed.
12. In PSK, the reference phase of the carrier must be ________________ at the receiver.
13. Changing phase only at a change of logic state ________________ bandwidth.
14. Multiplexing means the use of one facility to handle several separate but similar operations ________________.
15. With frequency division a guard ________________ is needed.
16. With time division a guard ________________ is needed.

ANSWERS

Fill in the Blanks

1. economical
2. narrow
3. noise
4. direct
5. signal to noise
6. amplitude
7. power
8. broad
9. reference level
10. half
11. fastest
12. reinserted
13. saves
14. simultaneously
15. band
16. time

CHAPTER 40

CARRIER SYSTEMS

The transmission of speech over long trunk routes is operated on a multi–channel carrier basis employing frequency division multiplex. The usable frequency spectrum of the line or cable is split into a number of adjacent bands, each of which is then employed as a separate channel. Modern systems are usually operated on a four-wire basis, which comprises a transmit channel and a receive channel, the two channels then form one circuit. Of all the systems in current use, the coaxial cable system is most important, as it forms the basic system for national network in all industrial countries. Twelve and twenty four channel systems are so arranged that they may form part of the coaxial cable system.

CARRIER SYSTEM

A speech signal can, by amplitude modulation, be changed in frequency from its *original* audio frequency (of 300–3400 Hz) up to a *higher carrier frequency*. Most of the world's long distance telephony systems utilise 12-channel groups: twelve voice channels are all changed in frequency in this way so that a *complete group of 12-channels occupies a 48 kHz bandwidth, basically from 60–108 kHz.*

Figure 40.1 is a block schematic of the *transmitting equipment* required for channels 1 and 2 of a standard 12-channel group. The audio input signal to a channel is applied to a *balanced modulator* together with the carrier frequency appropriate to that channel. The input attenuator ensures that the carrier voltage is 14 dB higher, than the modulating signal voltage, as required for correct operation of the modulator. The output of the modulator consists of the upper and lower sideband products of the amplitude modulation process together with a number of *unwanted components*.

Following the modulator is another *attenuator* whose purpose is two-fold. Firstly, it ensures that the following *bandpass filter* is fed from a constant-impedance source—necessary condition for optimum filter performance—and secondly, it enables the channel output level to be adjusted to the same value as that of each of the other channels. The filter selects the *lower sideband component* of the modulator output, suppressing all other components. To obtain the required selectivity, channel filters utilising piezoelectric crystals are employed. The outputs of all the twelve channels are *combined and fed* to the output terminals of the group. The *transmitted bandwidth* is 60.6–107.7 kHz, or approximately 66–108 kHz.

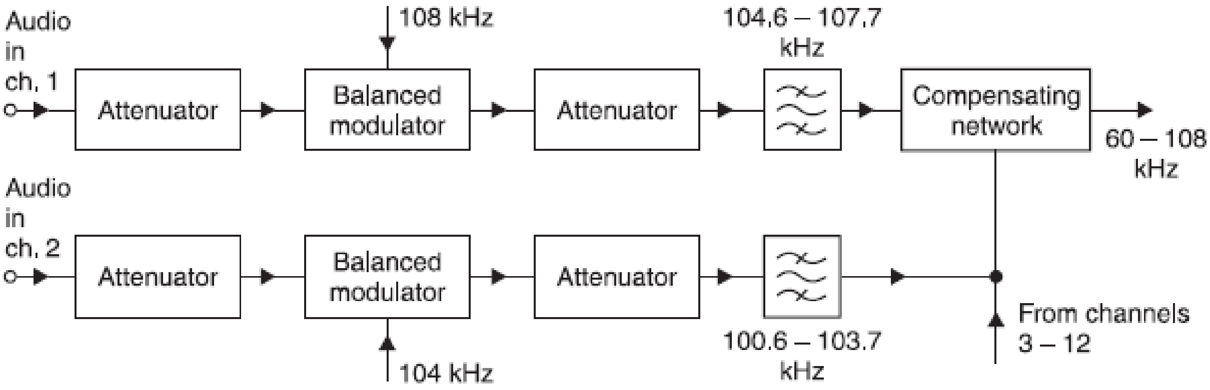


Fig. 40.1 Block schematic of the transmitting equipment required for channels 1 and 2 of a standard 12-channel group.

The equipment appropriate to channels 1 and 2 at the *receiving end* of the 12-channel group is shown in Fig. 40.2. The *composite signal* received from the line, occupying the band 60–108 kHz, is applied to the *twelve paralleled, channel filters.* Each filter selects the band of frequencies appropriate to its channel, 104.6–107.7 for channel 1, and passes it to the channel demodulator. The *attenuator* between the filter and the demodulator ensures that the filter works into a load of constant impedance. The demodulator is supplied with the same carrier frequency as that suppressed in the transmitting equipment. The lower sideband output of the demodulator is the required audio-frequency band of 300–3400 Hz and is selected by the *low-pass filter.* The audio signal is then amplified and its level adjusted by means of the output *attenuator.*

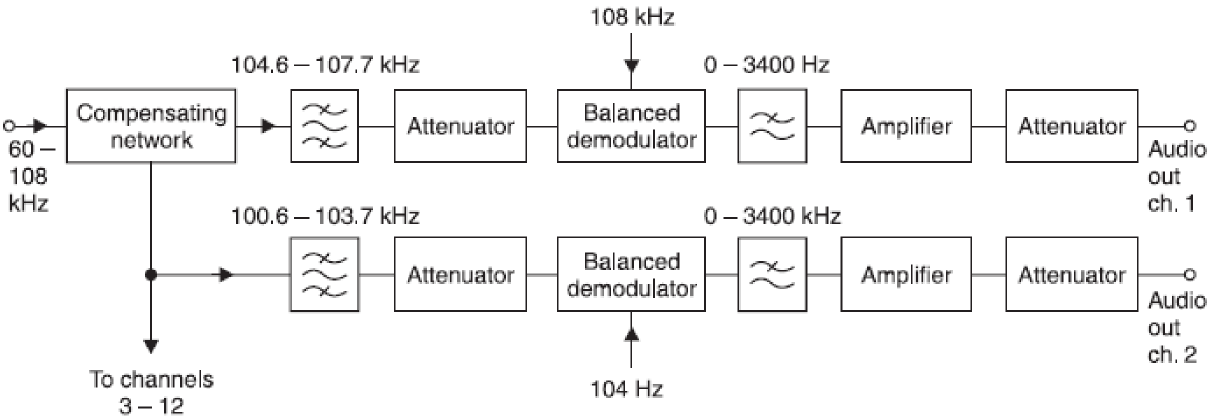


Fig. 40.2 Receiving equipment for a carrier system

The assembly of the basic 12-channel carrier group can be illustrated by means of a frequency spectrum diagram. The *spectrum diagram* of a single channel is given in Fig. 40.3. The actual speech bandwidth provided is 300–3400 Hz but a 0–4000 Hz bandwidth must be allocated per channel to allow a *900 Hz spacing* between each channel for filter selectively to build up.

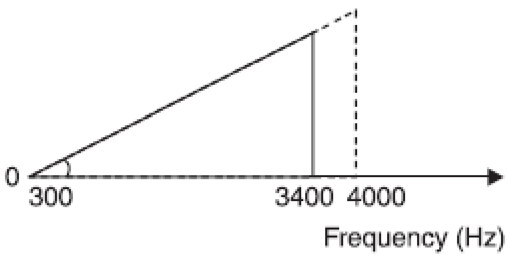


Fig. 40.3 Bandwidth for a commercial speech circuit

Figure 40.4 (a) shows the frequency spectrum diagram for the 12-channels forming a group; the carrier frequency of each channel is given and so are the maximum and minimum frequencies transmitted. It can be seen that *all the channels are inverted;* that is, the *lowest frequency* in each channel corresponds to the *highest frequency* in its associated audio channel, and vice versa. Since all the channels are inverted, the group may be represented by a single triangle as shown in Fig. 40.4 (b).

The 12-channel system can be used as a *building block* for the next larger *assembly stage* or as a system which can be transmitted to line in its own right. Five 12-channel groups can be combined to form a 60-channel supergroup. Five supergroups make up a 300-channel *mastergroup.* Three mastergroups make up a 3872 kHz bandwidth *supermaster* *group.* Alternatively, fifteen supergroups may be assembled direct to form a hyper group, sometimes called a 15 supergroup assembly.

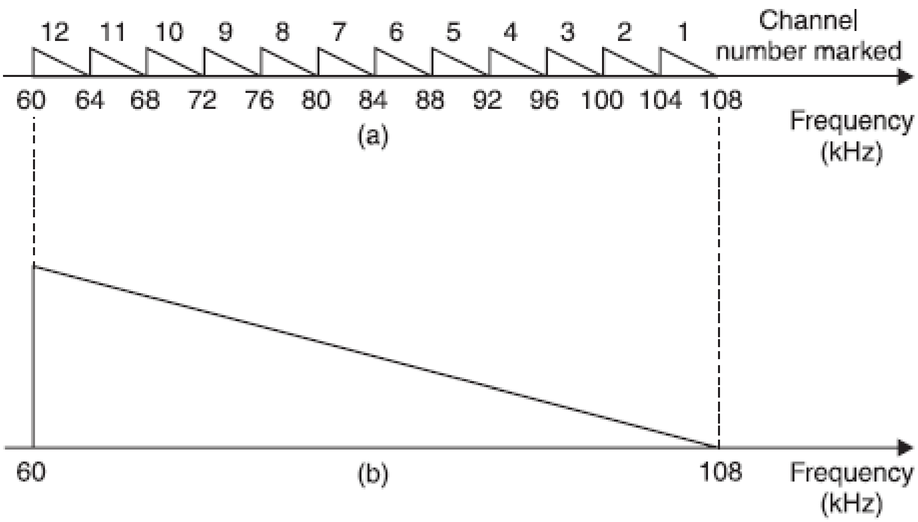


Fig. 40.4 Frequency spectrum diagram for a 12-channel group

SUBMARINE CABLES

Early submarine cables were all for *telegraphy*, the distortionless wider band needed for *telephony* could not economically be transmitted over long lengths of cable until component designers had produced *long-life trouble-free units*, capable of operating for many years in the depths of the oceans without fault incidence. These units, called *repeaters*, were designed to *amplify* incoming signals to offset the transmission losses incurred in the previous section of cable. It was also necessary to provide *equalisation* to offset the way different frequencies in the wideband information signal had been attenuated by different amounts. Such items were produced in the USA in the 1950s, and repeaters suitable for insertion in deep-sea cables were soon designed in America, Britain, France and Germany.

The first major submarine telephone cable was TAT-1. It had a capacity of 50 voice circuits. It is no longer in use. With the advance of technology it has become possible for wider bandwidths, meaning more circuits, to be transmitted. The latest *transatlantic cable*, TAT-6, has a 30 MHz bandwidth and an installed capacity of 4000–3 kHz circuits.

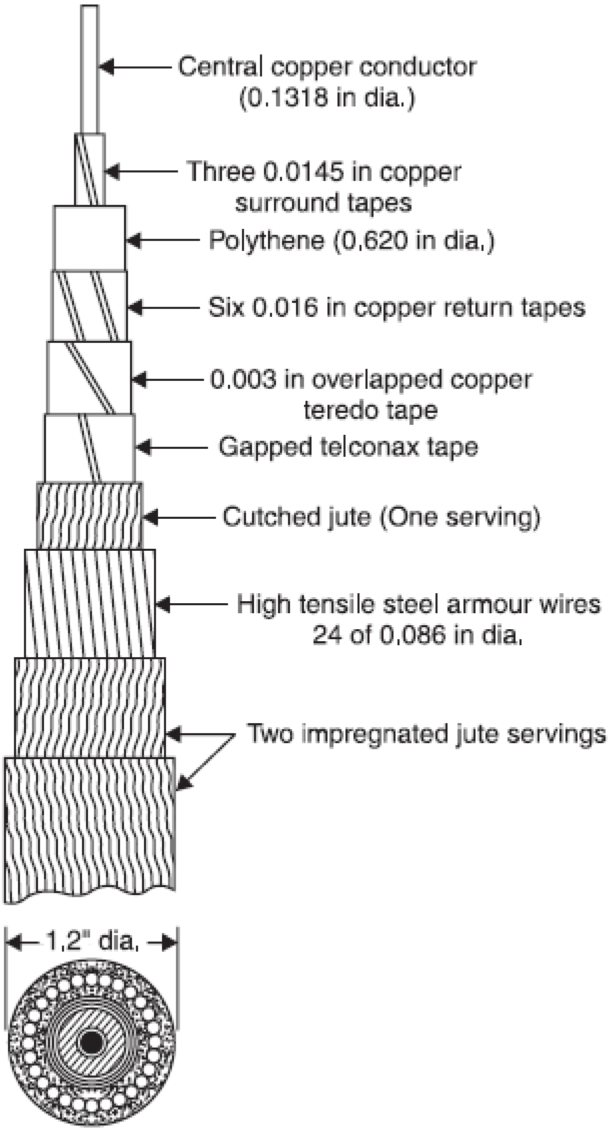


Fig. 40.5 Transatlantic telephone cable

A submarine cable is sometimes under considerable tension, especially when it is being picked up from the sea bed, so great tensile strength is necessary in addition to the ability of the cable and repeaters to withstand the high pressures of deep waters. Early telephone cables were made up in a generally similar way to the 19th century submarine telegraph cables, with one or more layers of heavy steel armour wire to provide the necessary protection (Fig. 40.5). The *cores* of the two types of cable were, of course, quite different; *telegraph cables* usually had a heavy, well insulated copper conductor to carry the signal current, with the return current normally flowing through the sea itself, where as *wide-band telephone cables* are of coaxial type, i.e. they have a central copper conductor, surrounded by a carefully dimensioned insulant, then copper tapes as a return-path outer conductor, then more insulant, then the outer armour wires.

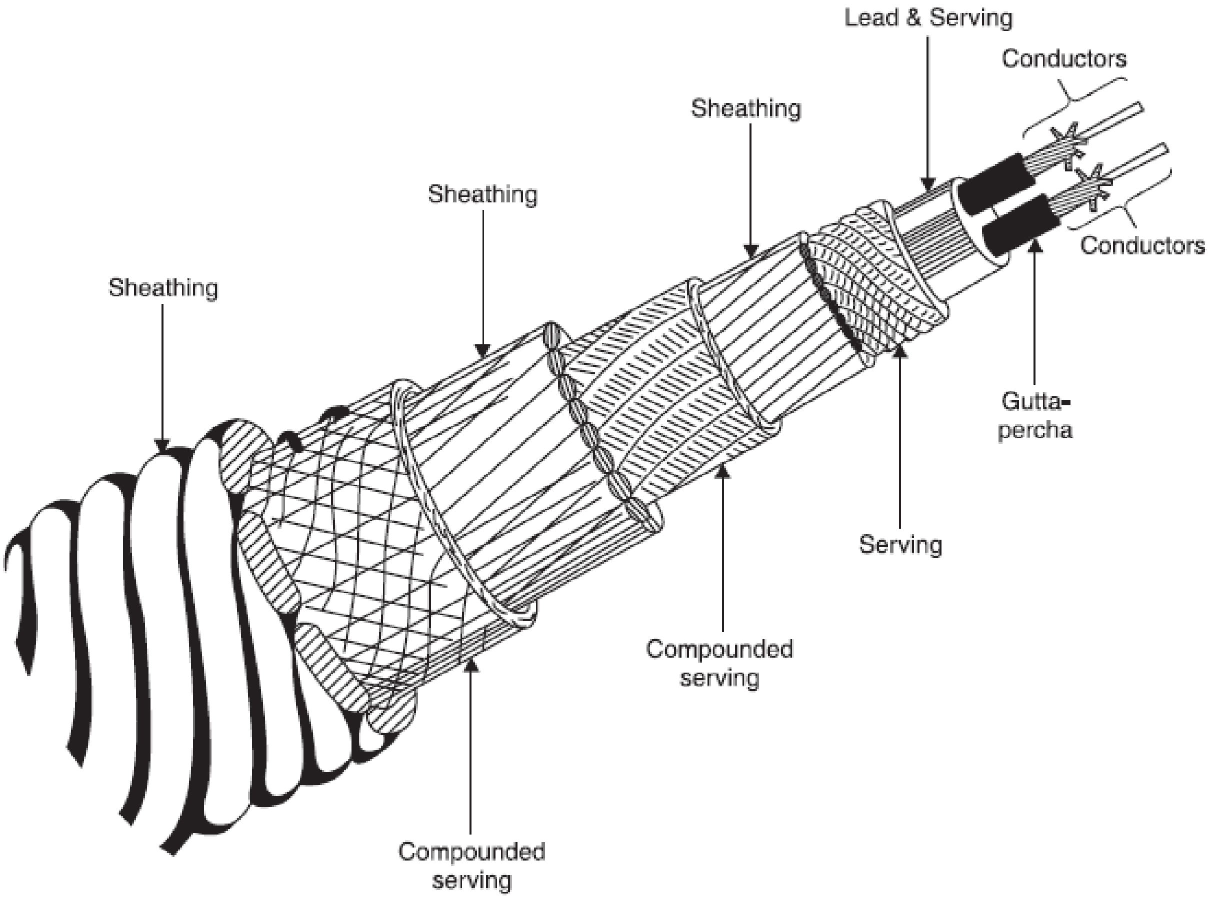


Fig. 40.6 Heavily armoured twin-core shore end telegraph cable

Cable systems are now used together with satellite radio systems. There is an economic and security need for both types of communication media to be used, most countries feel that it would be unwise to depend completely on any one system and no doubt *fibre optics* will enable cable manufacturer to provide circuits more economically than is now possible.

EXERCISES

Descriptive Questions

1. Draw the block diagram of the transmitting equipment of a carrier system and explain briefly.
2. Draw the block diagram of the receiving equipment of a carrier system and explain briefly.
3. Sketch the bandwidth for a commercial speech circuit.
4. Draw the frequency spectrum diagram for a 12-channel group.
5. What is the need for submarine cables?
6. Write short notes on :

    a. Supergroup

b. Mastergroup
c. Super master group
d. Hypergroup

Fill in the Blanks

1. Most of the world's long distance telephony systems employ ________________ groups.
2. A complete group of 12-channels occupies a bandwidth from ________________ to ________________
3. The attenuator ensures that the filter works into a load of ________________ impedance.
4. Repeaters are designed to amplify incoming signals to offset ________________.
5. The 12-channel group can be used as a ________________ for the next larger system.
6. The latest transatlantic cable has a 30 MHz bandwidth and an installed capacity of ________________ circuits.

ANSWERS

Fill in the Blanks

1.	12-channel
2.	60 kHz, 108 kHz
3.	constant
4.	transmission losses
5.	building block
6.	4000–3 kHz.

CHAPTER 41

FIBRE OPTICS

During the last hundred years a series of inventions have enabled telecomunication links to be built with ever-increasing capacity. The early telegraph links carried signals at speeds up to about 30 words per minute, or about 15 bits per second. Some of the fibre-optic cable being installed today in the longdistance trunk networks has the capacity to carry signals at a speed of 2.4 Gbps. The theoretical capacity of these hair-thin glass fibres is tremendous. Much fibre optic cable is being installed today, mainly in the long distance and inter-office networks. Over the next ten years, fibre optical cable will be progressively installed as local loops in some areas. In the meantime, however, a multitude of different transmisssion media will continue to be used. It will take some time before all coaxial cable and copper wire systems have been replaced with fibre.

THE TELEPHONE NETWORK

The telephone network is *the world's most complicated machine*. When a telephone call is made it may travel over many different types of channels, and *complex switching facilities* are needed to set up its path. Telecommunications systems can be divided into four main parts.

1. Instruments : The term instrument is used for *the device the subscriber employs to originate and receive signals*. The vast majority of instruments are *telephone hand sets*. Today, however, an endless array of other devices are being *attached* to telephone lines including computer terminals used for data transmission.

2. Local loops : These are the *cables that enter the subscriber's premises*. On a telephone network they connect the telephone handsets or other devices to the local switching office (central office). *Telephone loops today consist of wire pair cables*. Every subscriber has his own pair of wires to the local switching office, and nobody else uses it unless he is on a party line. There are several hundred million miles of telephone subscriber loops. *Coaxial cables* are also laid into homes by cable television organisations, and these have many potential uses other than television. *In the future, high capacity fibre optic cables will be used extensively in the local loop*.

If a local call is made, employing only local loops, to the telephone exchange, a larger bandwidth could be obtained, and a *faster data rate* could be transmitted. Some lengthy local loops have *loading coils* connected to them. These reduce attenuation but *lower* the potential bandwidth and bit rate that could be used. *To achieve high-speed transmission over the local loops, the loading coils must be removed*. Frequencies up to a megahertz can be transmitted over a telephone loop. To transmit such high-frequencies, *the signal levels must be carefully coordinated* to avoid interference with other services that may be using the higher frequencies on other pairs in the *same* cable. Generally the gain has to be *equalised* at short intervals. A baseband data transmission rate of 250,000 bps is generally attainable over *unloaded local loops*. As the local loop is progressively digitized loading coils become redundant.

The twisted-pair local loop is the weakest link and most costly portion of the telecommunications network. The term weakest link refers to the inability of the local loop to handle high-frequency signals.

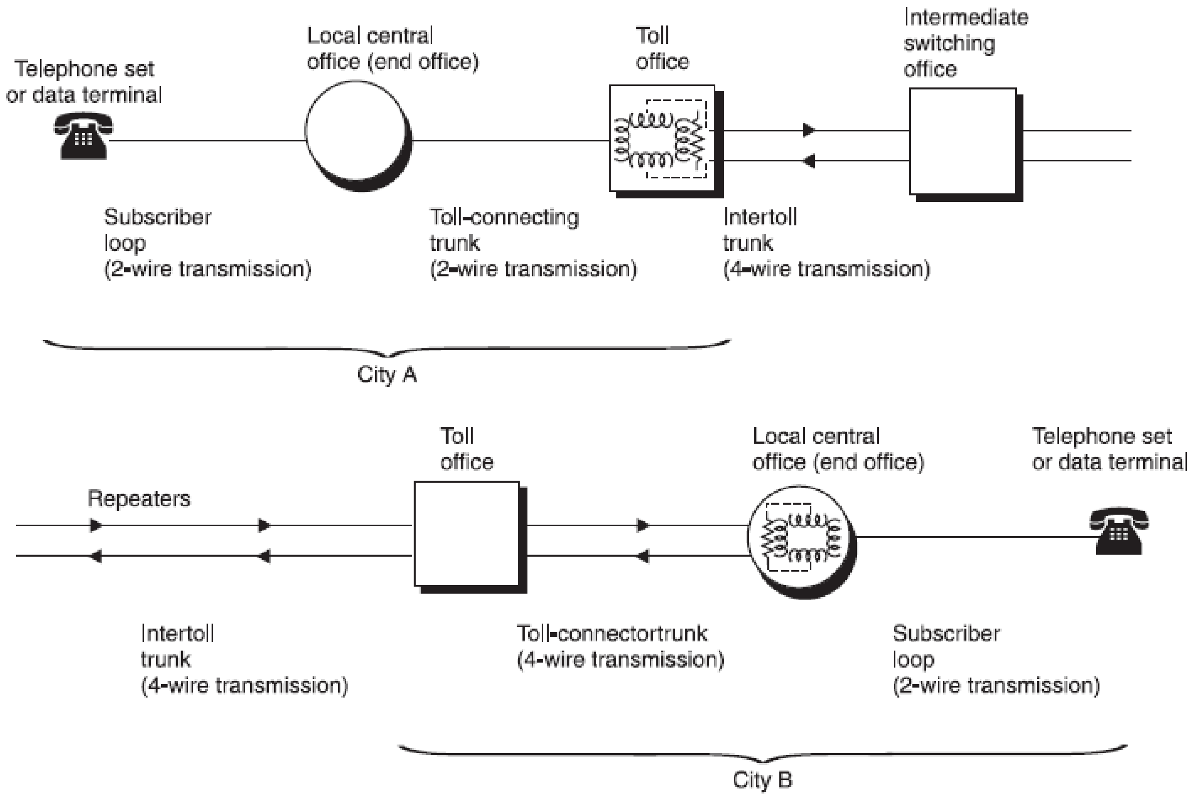


Fig. 41.1 An intercity telephone connection showing 2-wire and 4-wire paths. Most channels except for the subscriber loop are now digital.

The twisted-pair local loop has a narrow bandwidth. The other portions of the PSTN are composed of *wide-bandwidth facilities*. Voice and data travel across the PSTN on ribbons of fibre which can handle many high-bandwidth signals only to arrive at the local *loop* which *restricts* communication to low-frequency analog signals. The local loop is usually from a few hundred feet to several miles in length but it is often referred to as *the last mile*.

It is technically cost prohibitive to replace the existing wire cables with fibre optic cable and *compatible phones* in the local loop. To use fibre optic in the local loop, an *interface device* will be required at every phone to convert electrical energy into light waves. The device could be incorporated in the phone but this would require customers to buy new phones. Twisted-pair copper wire has provided us with a low cost medium to connect telephones to a central exchange and *the twisted pair local loop has served as an excellent transmission medium for analog voice signals*.

Increasing demand for the ability to place high-speed data signals over the local loop will eventually lead to the deployment of fibre optic facilities to replace *all, or a portion of* the local loop. The *first phase* of fibre introduction into the local loop has to use fibre as a main feeder cable to serve a segment of exchange territory. *Multiple subscriber line carrier systems* (SLC-96s in Fig 41.2) are placed on a multiplexer in the central office. Each SLC-96 connects to 96 line circuits. Demultiplexers are placed at several locations along the fibre route, and several SLC-96 field units are placed on the demultiplexer.

3. Switching facilities : An *elaborate network* of switching offices enables any telephone to be connected to almost every other telephone. Most of the *switching and control functions* are carried out entirely by computers. Some electromechanical facilities will continue to exist in less populated areas for a few years.

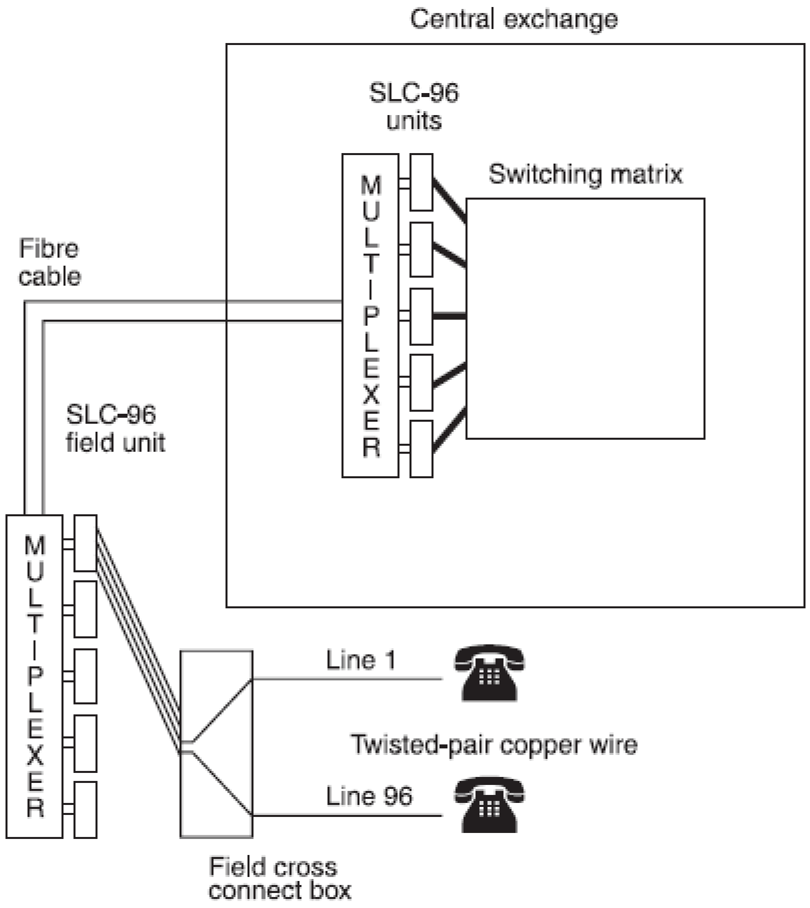


Fig. 41.2 SLC-96 units over a fibre

4. Trunk Circuits : Trunk circuits are *transmission links which interconnect the switching offices*. Such links normally carry more than one telephone call. On high-traffic routes, they carry many thousands of calls *simultaneously*. A variety of different *transmission media* is employed including wire pairs, coaxial cables, microwave radio, satellites, and fibre optics.

NONVOICE TRAFFIC

For half a century *telephone technology* has dominated telecommunications. There have been 1000 times as many telephone subscribers as other types of telecommunication users. *Consequent economics of scale have dictated that telegraph and data traffic should be converted to a form in which they can travel over the telephone system.*

Since 1970s new types of *common carriers* emerged and built their own *non telephone networks*. Separate data transmission networks for computer users are available in all major industrial countries. Some of these operate by attaching special equipment to the telephone networks. Others employ new transmission networks, physically separate from the *telephone networks*.

Telephone traffic and computer traffic have *characteristics* so different that different network architectures are needed.

TELEPHONE USERS

1. Require a fixed capacity channel.
2. Always carry out a a two-way conversation.
3. Tolerant of noise on the channel.
4. Transmit or listen continuously until the call is disconnected.
5. Require immediate delivery of the signal.
6. The transmission rate is constant.
7. The time to set up the connection can range from a few seconds to 1 minute (maximum).

COMPUTERS AND THEIR TERMINALS

1. Require a very wide spread of channel capacities ranging from a few bits per second to millions of bits per second.
2. One-way or two-way transmission.
3. Data must be delivered without errors.
4. In a man-computer dialog, transmission is in bursts.
5. In non-real-time data transmission the data can be delivered later, when convenient.
6. In a man-computer dialog the mean number of bits per second is usually low, but the peak requirement is often high. The peak-to-average ratio is often as high as 1000.
7. It is desirable that the connection be set up in a second or less.

Computer users have fundamentally *different* characteristics and requirements to plain old telephone service. Table 41.1 summarises the differences. An important difference is the *peak-to-average transmission requirement*. It is 1 in telephone conversations but can be 1000 or even higher in human-computer dialogs. *The higher the peak/average ratio, the greater the inefficiency of using a transmission channel which transmits at a constant fixed rate. Instead a channel is needed which transmits bursts of data when they are required with a suitable short response time*. Such a channel can be derived either by using suitable equipment to share a fixed-rate channel or by using a common carrier network architecture which allocates capacity on a rapidly *time-varying basis*.

FIBRE IN LOCAL LOOP

Basically, there are four ways to reach the local telephone company's end office, that is, there are only four ways to implement the local loop): copper wire, coaxial cable, fibre optic cable, and wireless communication. The copper is owned by the local phone company and they try to cram as much information as is technologically possible. The *coaxial cable* is owned by the local cable company. They too are trying to cram as much information down the cable as is technologically possible. Today, running *fibre optic cable* to individual homes is considered infeasible. And then there is *wireless*.

To date the majority of *optical-fibre-systems* have been installed in the trunk and inter-office network. The next stage is the introduction of fibre optics into the local loop, taking a fibre cable right into a subscriber's office or home. For many *new* local subscriber networks, it would be cheaper to instal fibre cable than copper. By using *wave-division multiplexing* techniques both the incoming and outgoing signals can be multiplexed onto a single fibre; so only one fibre is needed per household. The potential of having fibre in local loop is enormous. The 565 Mbps systems currently operating in the *trunk network* can easily be applied to the local loop. A *single 565 Mbps cable* into the home could, for example, provide the following:

- Eleven simultaneous 50 Mbps high-definition television (HDTV) channels with compression
- Eleven 50 Mbps data channels
- Numerous channels for; videophones; music; ordinary voice channels; telemetry; photographic-quality colour facsimile television and video wall screens; access to vast music and video libraries.

The nature of the telephone network is such that *improvements in bandwidth tend to be implemented first in the trunk network, and they gradually cascade down to the subscriber* as the cost of new technology reaches a point where implementing into offices and, later, homes of millions of subscribers becomes economical.

In future, optical technology in the form of transmission and switching will dominate telecommunications. The impact this will have on how and what we communicate is likely to be as revolutionary as was the invention of the telephone itself.

OPTICAL SYSTEMS

Although many lasers and LEDs are able to produce outputs in the visible light band, most current optic fibre telecommunication systems (Fig. 41.3) use signals of wavelengths 0.8 μm to 1.3 μm, both in the *infrared band*. These are still called optical systems: even though the signals cannot be seen they are transmitted in exactly the same way as visible light signals. The main reason why engineers wanted to be able to modulate *coherent light* was to take advantage of the *tremendous bandwidths* which could be carried by these very high-frequencies. The figures below are typical figures, indicating the orders of magnitude involved.

Table 41.1 Carrier Wave

	Frequency	Wavelength	Possible bandwidth per system
HF radio	3 MHz	30 m	16 kHz (4 voice channels)
Microwave radio	6 GHz	5 cm (10^{-2}m)	4 MHz (960 voice channels)
Optic fibre	*100000 GHz*	*1 μm* (10^{-6}m)	*Several thousand MHz Hundreds of thousands of voice channels—but only a few thousand are possible with current technology*

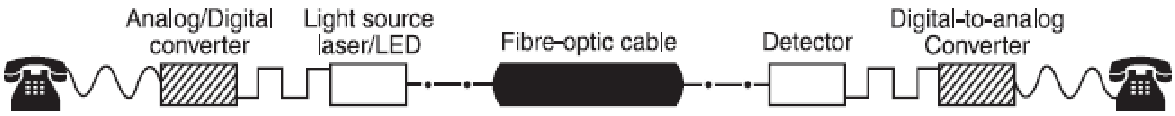


Fig. 41.3 Basic principles of fibre-optic communication

OPTICAL FIBRE CABLE

An optical fibre cable consists of a glass *core* that is completely surrounded by glass *cladding* (Fig. 41.4). The core performs the function of *transmitting* the light waves, while the purpose of the cladding is to minimise surface losses and to *guide* the light waves. The glass used for both the core and the cladding must be of very high purity since *any impurities present will cause some scattering of light to occur*. Two types of glass are commonly employed : *silica-based glass* (silica with some added oxide) and *multi component glass* (e.g. sodium borosilicate).

Some new optic fibres do not use glass at all, but special types of plastic; these are usually cheaper to make than very pure glass but introduce greater attenuation. Fibres now being manufactured are so free from impurities that very little energy is lost as the signals travel along. *An attenuation of less than 1 dB per kilometre is not uncommon for the latest high-purity silicons*.

A major constraint with optic fibres (apart from the straight forward one of attenuation) is that since the wavelength of light is very short a light wave signal injected into one end of an optic fibre (sometimes called an *optic wave-guide)* does not merely travel straight down the middle of the core. It *swings* from side to side, continually being *reflected or refracted* back from the core/cladding surface. Clearly a pulse going straight down the middle will reach the end just before those parts of the same pulse signal which have *zigzagged* along, taking a longer path. This places a restriction on the *maximum possible bit rate* that may be transmitted satisfactorily in the fibre.

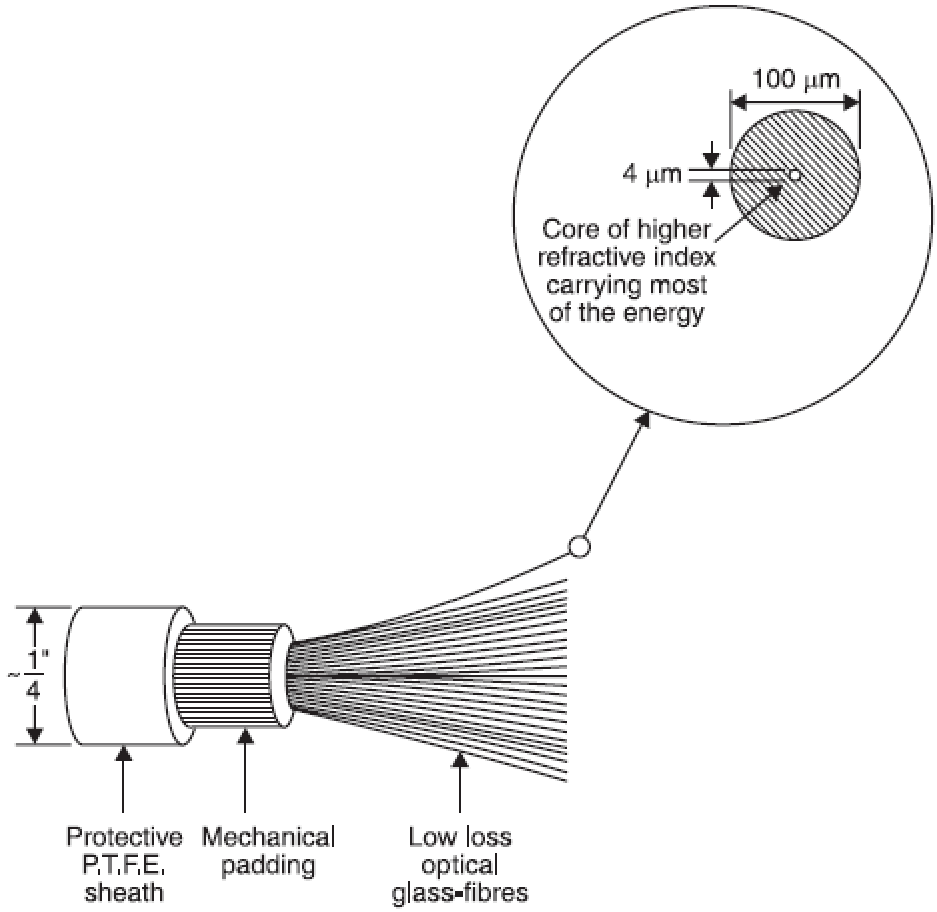


Fig. 41.4 Optical fibre cable containing single mode fibres

TYPES OF OPTICAL FIBRE

It would have been ideal if right from the beginning we could have used fibres made with tiny diameters *comparable with* the wavelength of the optical signal being used, so that no zigzagging was able to take place, but the manufacture and joining of such high-precision fibres presents considerable difficulty. Today's more common thicker-core fibres are called *multimode* (because many different modes of transmission are possible). Fibres with very small diameter cores are called *monomode* (because only a single mode of transmission is possible). There are three basic types of optial fibre:

1. Stepped-index multimode: The basic *construction* of a stepped-index multimode optical fibre is shown in Fig. 41.5 (a) and its *refractive index profile* is shown in Fig. 41.5 (b). It can be seen that *an abrupt change in refractive index of the fibre occurs at the core/cladding boundary*. The core diameter, $2r_1$, is usually 50–60 μm but in some cases may be up to about 200 μm. The diameter $2r_2$, of the cladding is *standardised* whenever possible at 125 μm.

> *Stepped-index multimode fibre produces large transit time dispersion* (Fig. 41.6) so its use is restricted to applications such as those involving comparatively *low speed data signals*.

2. Stepped-index monomode: Figure 41.7 (a) shows the basic *construction* of a stepped index monomode optical fibre and Fig. 41.7 (b) shows its *refractive index profile*. Once again *the change in the refractive indices of the core and cladding is an abrupt one but now the* *dimensions of the core are much smaller*. The diameter of the core is of the same order of magnitude as the wavelength of the light to be propagated, it is therefore in the range 1–10 μm. The cladding diameter is the *standardised* figure of 125 μm.

Stepped-index monomode fibres are *at present difficult and expensive to manufacture and join*, so most of the stepped-index optic-fibre telecommunication systems now installed use multimode fibres. As the technological difficulties are overcome the use of monomode fibres may well increase.

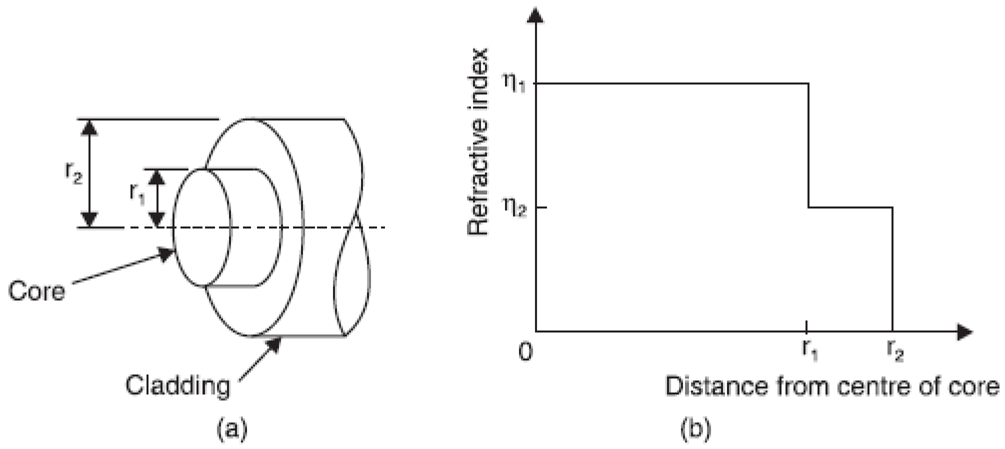


Fig. 41.5 Stepped index multimode optical fibre : (a) cladding and core (b) refractive index profile

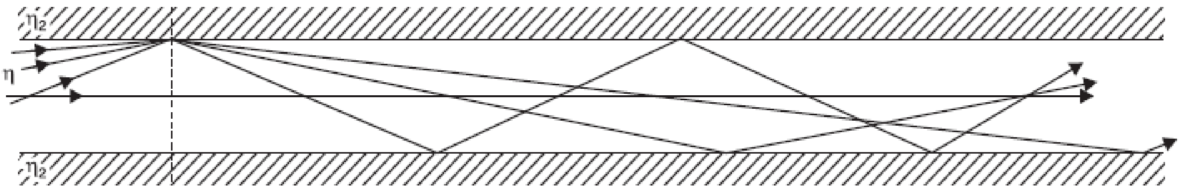


Fig. 41.6 Multimode propagation in a stepped-index fibre

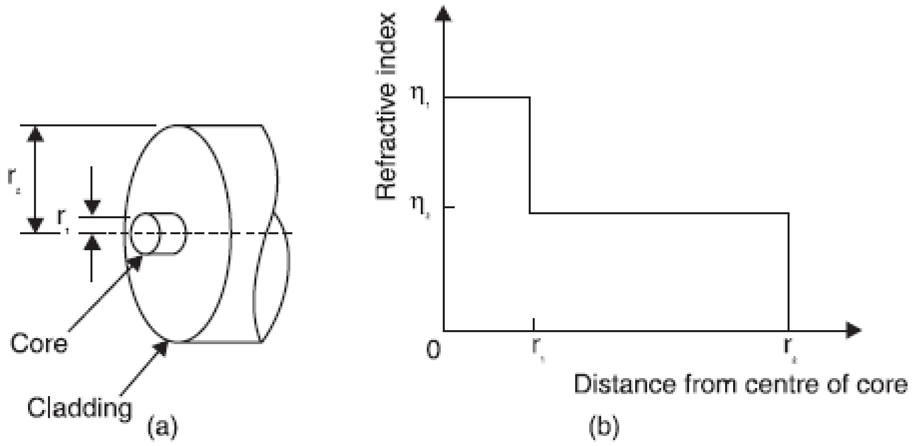


Fig. 41.7 Stepped-index monomode optical fibre (a) cladding and core (b) refractive index profile

3. Graded-index multimode: The basic construction of a graded index multimode optical fibre is the same as that of a stepped index multimode fibre shown in Fig. 41.5 (a). The core diameter is also in the range 50–60 μm and the cladding diameter is 125 μm. The refractive index of the inner region or core is the highest at the centre and then *decreases parabqlically* towards the edges, Fig. 41.9, to that of the cladding material. This means that *light waves will be refracted back from the outer boundary of the fibre, and not reflected as with stepped-index fibres* (see Fig. 41.10). So waves will go *straight down* the centre of the core or *zig-zag* from side to side as they do in stepped-index fibre, but in a *smoother manner*. The main difference is, however, that waves which zig-zag along in a graded index fibre pass through regions with a lower refractive index than in the central part of the core, so *although they travel a greater distance, it is at a higher velocity*. The effect of this is to reduce the differences in the times taken by the many different modes; ideally, *all modes then arrive at the distant end in exact synchronism*.

No matter which of the three possible types of propagation is used, *the dimensions of the outer medium or cladding must be at least several wavelengths*. Otherwise some light energy will be able to escape from the system and extra losses will be caused by any light scattering and/or absorbing objects in the vicinity.

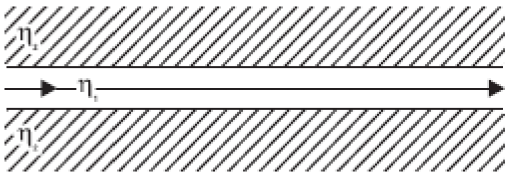


Fig. 41.8 Monomode propagation in a stepped-index fibre

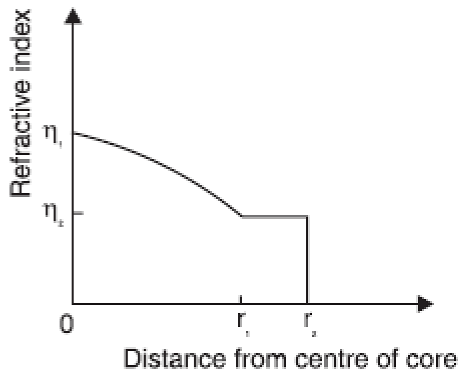


Fig. 41.9 Refractive index profile of a graded-index multimode optical fibre

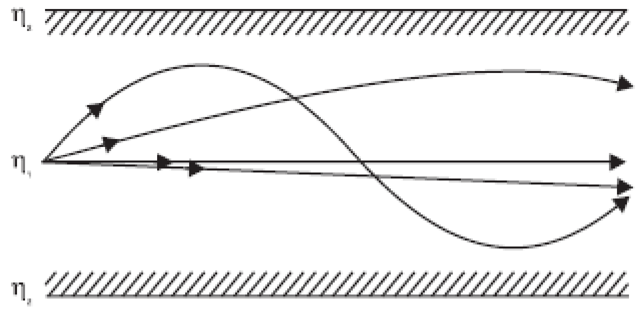


Fig. 41.10 Multimode propagation in a graded-index fibre

OPTICAL FIBRE ADVANTAGES AND DISADVANTAGES

Fibre optic cabling is immune to electrical interference. Signals are transmitted in the form of *'off and on' light pulses* similar to a flash light. No electricity is present in transmission over fibre. Thus signals carried on strands of fibre do not interfere with each other. Therefore fibre can be run in areas without regard to interference from electrical equipment.

Advantages of fibre cabling are:

1. *Low transmission loss:* This permits longer repeater sections than with coaxial cable systems, thereby reducing costs.
2. *Wide bandwidth:* This means a large channel carrying capacity.
3. *Small cable size and weight:* This means that drums of cable can be run economically and that each cable uses less space in cable ducts.
4. *Immunity to electromagnetic interference:* This permits use in noisy electrical environments, such as alongside electrified railway tracks and means that low signal-to-noise ratios are acceptable.
5. *Non inductive:* The fibre does not radiate energy so causes no interference to other circuits. Communication security is thereby enhanced.

6. *Long-term cost advantages:* The basic raw material silica is never likely to be in short supply and improved technology is continually producing lower cost and more efficient devices.

Disadvantages of fibre cabling are:

1. *Termination, component and connector costs:* These are higher than for copper wiring. Special equipment is required to terminate fibre cables within buildings, test and splice fibre and to convert electrical signals to light pulses and vice versa.
2. *More care in handling:* Fibre is not as flexible as twisted pair in bending around coners, therefore, more care in handling fibre is needed.
3. *Local electrical power is required:* When fibre is brought into buildings from telephone companies or to curb in residential areas local electrical power is required. This adds a point of vulnerability in the event of a power outage.
4. *Specialised technicians:* Who might be paid at higher levels, often are required to work with and test fibre cabling.

EXERCISES

Descriptive Questions

1. Telecommunication systems can be divided into four main parts. Explain
2. Compare telephone traffic with computer traffic.
3. Illustrate and explain the basic principles of fibre-optic communication.
4. What are the types of optic fibre?
5. What are the advantages and disadvantages of optical fibre?
6. Write short notes on the following:

    a. Fibre in local loop
    b. SLC-96
    c. Optical fibre cable

Fill in the Blanks

1. The term instrument is used for the device the subscriber employs to ____________________ signals.
2. Local loops are the cables that enter the ____________________ premises.
3. Loading coils reduce ____________________ but lower the potential bandwidth and bit rate that could be used.
4. To achieve high speed transmission over local loops, the loading coils must be ____________________
5. Generally the gain has to be equalised at ____________________
6. The twisted pair local loop is the ____________________ link.
7. The twisted pair local loop has a ____________________ bandwidth.
8. Most of the switching and control functions are carried out entirely by ____________________
9. Trunk circuits are ____________________ which interconnect the switching offices.

10. Telegraph and data traffic should be ____________________ to a form in which they can travel over the telephone system.
11. There are only ____________________ ways to implement a local loop.
12. Improvements in bandwidth in a telephone network tend to be implemented first in the ____________________ and then gradually cascade down to the subscriber.

ANSWERS

Fill in the Blanks

1.	originate and receive
2.	subscriber's
3.	attenuation
4.	removed
5.	short intervals
6.	weakest
7.	narrow
8.	computers
9.	transmission links
10.	converted
11.	four
12.	trunk network

CHAPTER 42

DATA SERVICES

For most of its past history, telecommunications has been dominated by analog switching, transmission and frequency–division multiplexing. Telecommunication, networks in the 1980s underwent a process of rapid change. The conventional analog telephone was replaced steadily by computer controlled digital networks of immense capacity. The conversion of all telecommunication networks to digital switching and transmission will take some considerable time. Until then, digital and analog facilities must coexist, particularly in many developing countries. Voice is no longer the dominant feature around which networks are designed. From now on networks should carry all types of signals in a digital fashion.

The key concept of a digital network is that it lays the foundation for creating an ubiquitous integrated network capable of handling all our communications requirements, from a simple telephone call to the broadcasting of high definition television.

WHY DIGITAL?

What are the *advantages* of transmitting the telephone voice in digital form? One of the major advantages existed on early teletype links, but when several teletype channels were multiplexed into one voice circuit, the advantage was lost. Now the changing economies are bringing it back. With *analog transmission*, whenever the signal is amplified, the noise is amplified alongwith it. As the signal passes through its many amplifying stations, the noise is cumulative. With *digital transmission*, however, each repeater station regenerates the pulses. New, clean pulses are reconstructed and sent on to the next repeater, where yet another *cleaning-up process* takes place. So the pulse train can travel through a *dispersive noisy medium* but, instead of becoming more and more distorted until eventually parts are unrecognizable, it is repeatedly reconstructed, thus *remaining impervious to most of the corrosion of the medium*. Of course, an exceptionally large noise impulse may destroy one or more pulses so that they cannot be reconstructed by the repeaters.

A major *disadvantage* of digital transmission would appear to be that *much greater bandwidth is required*. However, since the signal is reconstructed at frequent intervals down the line, it can tolerate much more battering than if it had to travel a long distance without reconstruction. It can survive traveling over a channel with a high level of distortion and with a poor signal-to-noise ratio. The trick that makes digital transmission worthwhile is *to reconstruct the signal repeatedly* so that it survives the bad distortion. A high bit rate can then be transmitted. This is shown in Fig. 42.1.

Consider a telephone wire pair under the streets of a city. With analog transmission it can carry a channel group-12 voice channels. Now suppose that we transmit *digitally* over the same wire pair. The digital signal becomes *distorted* as it is transmitted, as shown in Fig. 42.1. We catch it before it becomes too distorted to recognize whether a bit is 0 or 1. The bit stream is then *reconstructed, retimed and retransmitted*. The faster a bit stream is transmitted, the greater will be the distortion and the closer the *spacing* of repeaters necessary to reconstruct the signal correctly. How closer can they reasonably be spaced? There is a manhole or other access to the wires about every 600 feet. With that spacing, today's telephone wire pairs can be made to carry 2.048 Mbps, which is equivalent to 32 telephone speech channels.

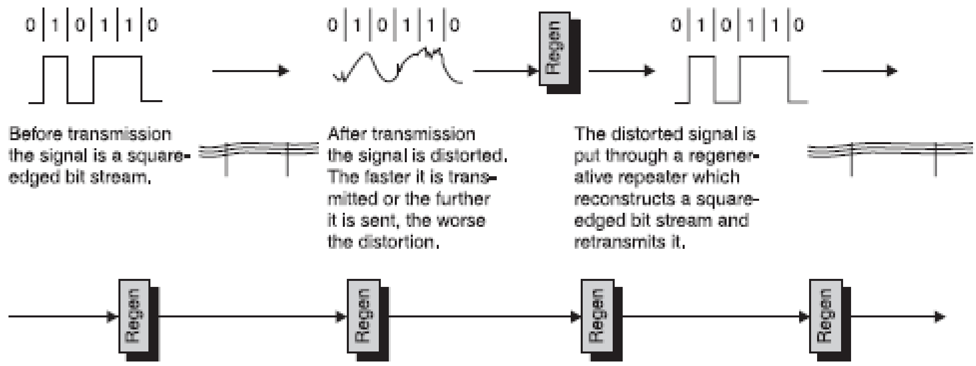


Fig. 42.1 A digital transmission line has many regenerative repeaters. If they are at sufficiently frequent intervals, a much higher bit rate can be transmitted than with analog transmission.

DIGITAL EXCHANGES

It has been conventional to divide telephone exchanges into two main categories : space division (or analog), and time division (or digital).

1. *Space Division* (or Analog) – in which *direct physical paths* are established right through the exchange from one subscriber's line to another. Connections may be metallic contacts or by solid-state analog devices.

2. *Time Division* (or Digital) – in which some or all of the switching stages in the exchange *operate by shifting signals in time*. Basically a connection is made between incoming and outgoing channels by transferring each PCM word from the *time-slot* of the incoming channel to that of the outgoing channel. This *time-shifting* is carried out by means of stores. Information is *written into* an address in a store, then during every cyclic scan the information from that particular address is *read out* so that it occupies the required outgoing time-slot.

Most digital exchanges are built from *subsystems*. These subsystems can be put together to provide a variety of exchanges for use at different points in the network (see Figs. 42.2 and 42.3). The simplification of function which follows the final elimination of FDM and analog trunks and junctions will be apparent. When this stage is reached, exchanges will take up a fraction of the floor area now occupied with great consequent savings.

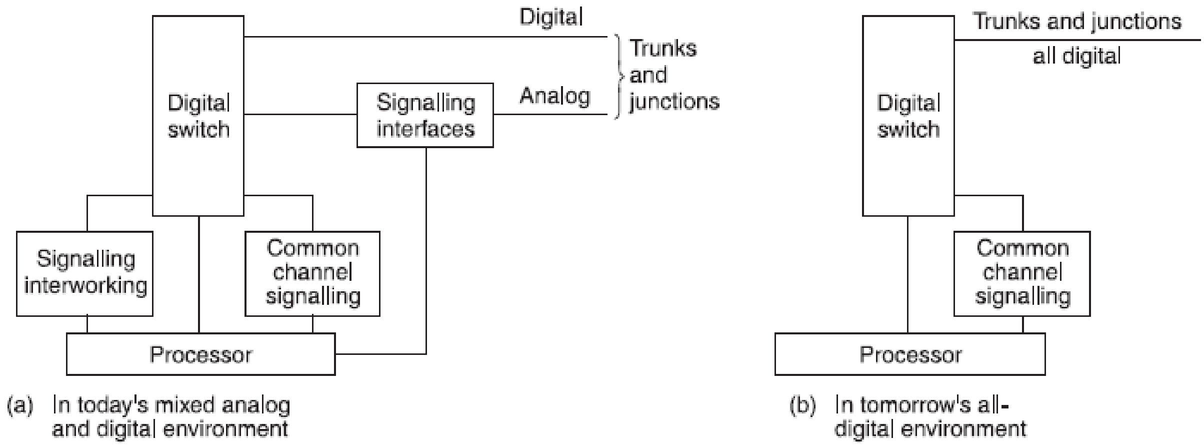


Fig. 42.2 Digital trunk exchange

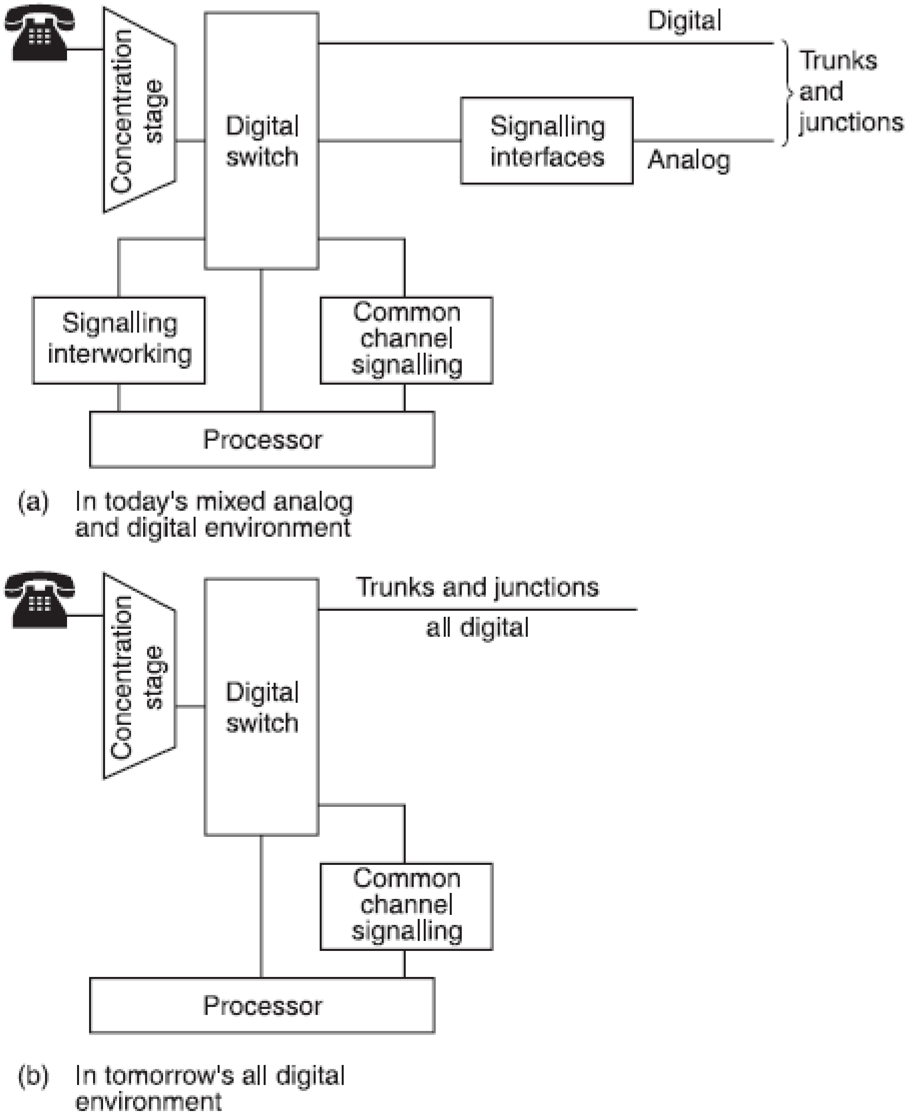


Fig. 42.3 Digital local exchange

The function of the *concentration stage* is to interconnect subscriber lines with the main switching subsystem in the exchange, the digital or group switching stage. The concentration stage has to be carefully designed to provide the grade of service required by the administration, with the greatest possible economy. Business lines are used much more than residential lines, so it is not possible to concentrate business lines to the same extent as *low calling rate* residential lines. Junctions to other exchanges are not concentrated at all; every trunk and junction circuit always has direct access to the group switch.

When a concentration stage serves the called subscriber, it is sometimes called the *expansion stage*, and shown separately even though it is sometimes the same piece of equipment as the concentration stage.

THE BORSCHT FUNCTIONS

The *interface* with the subscriber's line is at present the costliest part of all digital exchanges, largely because a 10000 lines exchange has to have 10000 of these units, each with all the features needed to *inter work* with various types of subscribers' lines. It is customary to describe these features as *the BORSCHT functions*, based on the initials of the key words.

1. *Battery feed to line* – there is normally no active power source at the subscribers' telephone; all the power, needed to drive the microphone and keypad, is fed out from the exchange *along the subscriber's line*.
2. *Overvoltage protection* – solid-state devices are very sensitive to high voltages, and *rapid-action protective devices* have to be provided in each line circuit so that if lightning does happen to strike an external line the exchange will not be put out of action.
3. *Ringing current injection and ring trip detection* – the bell at the called subscriber's telephone has to be rung; this means that quite a high voltage. an ac signal (sometimes about 70 volts) has to be *connected* to the line, to ring the bell; and as soon as the handset has been picked up off its cradle (gone *off hook)* the ringing must be *tripped and disconnected*.
4. *Supervision of the line* – equipment has to be provided which continually monitors the line so that as soon as it goes *off hook* (and a continuous dc path provided through the instrument) the exchange connection is *activated* and dial tone sent out to the caller. Dial pulses represent breaks in the continuous dc loop; these have to be *detected and counted* so that the exchange knows what number is required. When the caller finally clears down, by going back *on hook* the exchange must *break down* the established call and *note the time* at which this has been done for charging purposes.
5. *Codec (short for encoder plus decoder)* – this turns the analog signal received from the telephone instrument into a *digital signal* ready to be multiplexed with others, into a PCM signal. Incoming signals are similarly decoded from digital to *analog* before being sent out to the subscriber's instrument. Some digital exchanges have one codec per line; some *share* codecs between several lines; some use even fewer codecs by placing them between their concentration stages and the links going to their group switching stages.
6. *Hybrid for 2-wire to 4-wire conversion* – the line to an ordinary telephone subscriber uses one pair of wires, for both directions of conversation; this is called a *2-wire circuit*. The circuits inside a digital exchange use two electrically separate paths, one for each direction of transmission; this is called a *4-wire circuit*. To join together a 2-wire and a 4-wire circuit, a special device called a *hybrid coil* is used; this allows speech from the 2-wire line to enter the *4-wire transmit path*, and speech from the *4-wire receive path* to 2-wire line, but it blocks speech incoming on the 4-wire receive path from going out again on the 4-wire transmit path of the same circuit. Ordinary electromechanical exchanges use 2-wire circuits all the way through so there is no need for hybrids in these exchanges.
7. *Testing of both line and equipment* – it is necessary to be able to *test* the subscriber's line electrically so that faults may be *located and cleared*.

Any form of time-division switching is a one-way function, to provide a single *bi-directional* speech circuit through an exchange, two-channels, therefore, have to be switched through, one for each direction of transmission. Figure 42.4 shows how this is done in some exchanges.

LOCAL DISTRIBUTION NETWORKS

The connection between the subscriber and local telephone exchange consists of a pair of wires in a telephone cable. Since a large telephone exchange may have 10000 or more subscribers, the *local line network* can be quite complicated, particularly because provision must be made for *fluctuating demand*. The local line network is provided on the basis of forecasts made of the *future demand* for telephone service, *the object being to provide service on demand and as economically as possible*.

Since the demand fluctuates considerably there is the problem of forecasting requirements and deciding how much plant should be provided *initially* and how much at *future* dates. No matter how carefully the forecasting is carried out, some errors always occur and allowance for this must be made in the planning and provision of cable, i.e. *the local network must be flexible*. A network must be laid out so that the situation should not arise where potential subscribers cannot be given service in some parts of the exchange area while in other parts spare cable pairs remain.

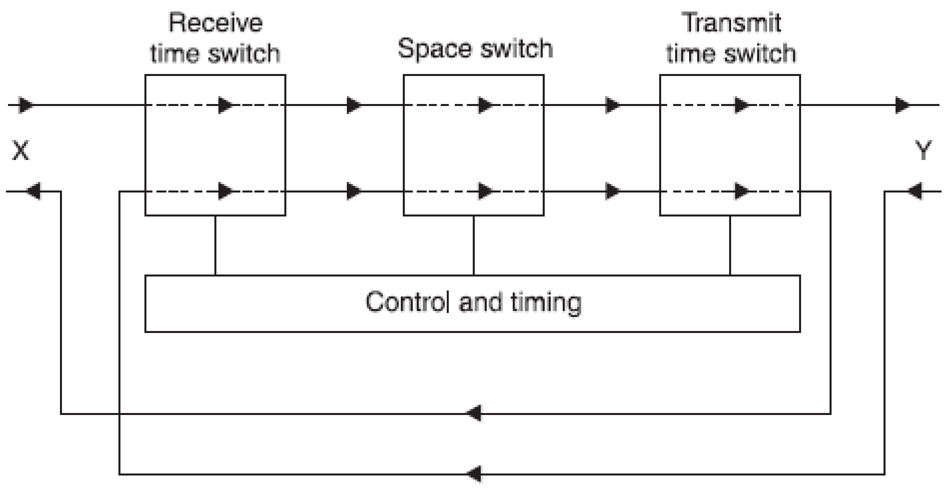


Fig. 42.4 Time-space time switching

The modern way of laying out a *local line network* is shown in Fig. 42.5. Each subscriber's telephone is connected to a *distribution point*, such as a terminal block on a pole or a wall. The distribution points are connected by small distribution or *secondary cables* to cabinets. *Primary cables* then connect these cabinets to the telephone exchange.

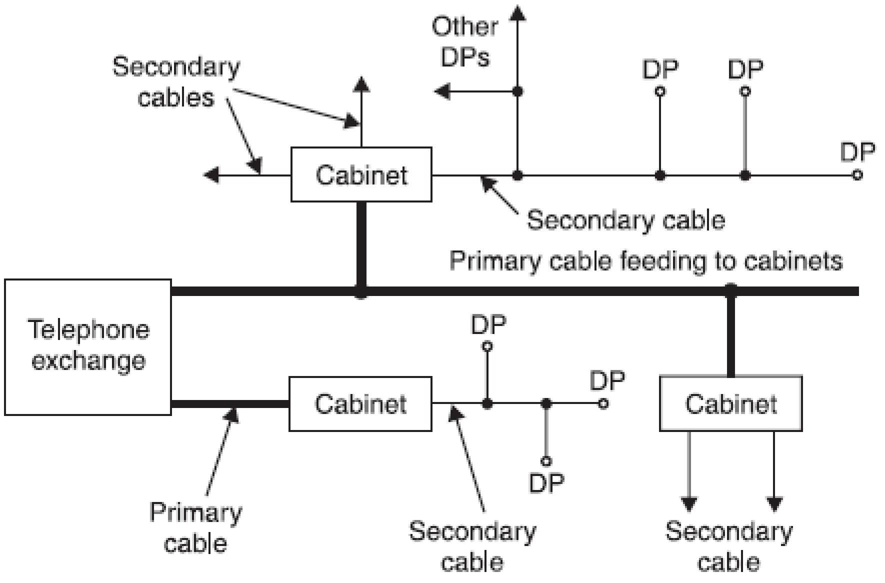


Fig. 42.5 Layout of a telephone exchange

It is usual to provide secondary cables on the basis of an *expected life* of about 15 years and the much larger primary cables for only about 5 years. This does not mean that these cables are expected to be scrapped after these periods; they are expected by then to be fully utilised and to require supplementing by *additional* cables. Cabinets are sometimes called *flexibility points:* if demand is much heavier than forecast in one part of the area served by the cabinet, and less in another part, cable pairs may be connected through at the cabinet to the faster developing area.

It is usually economically desirable to use underground ducted cable from exchange to flexibility points (cabinets) and then either go underground or overhead to distribution points depending on circumstances. *A ducted system*, although initially more costly than a *direct buried system* provides much greater flexibility for the future installation of additional cables, and facilitates cable repairs or replacement.

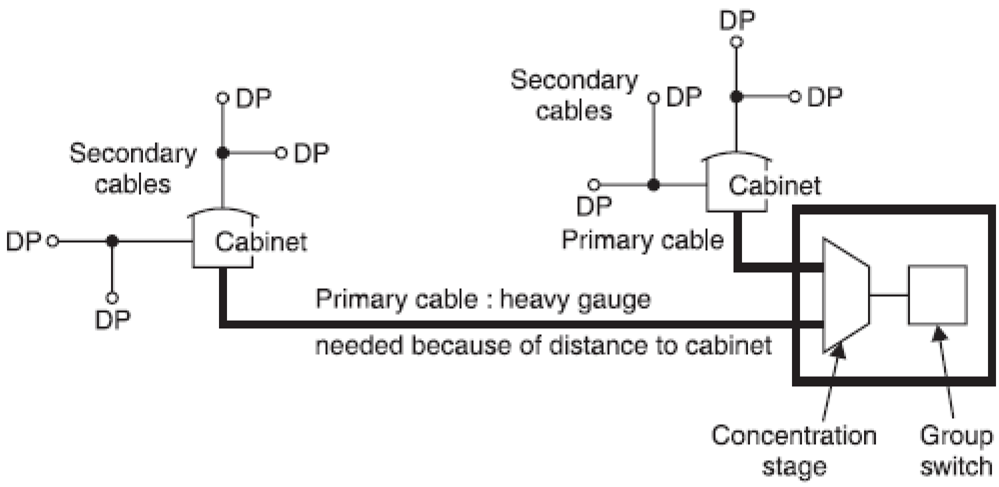


Fig. 42.6 Traditional-type cable distribution

At the present time, *polyethylene-insulated,* conductors and *polyethylene-sheathed* cable incorporating an aluminium screen and water vapour barrier is the most economical type of cable to be used in ducts. *Main or primary* cables from exchanges to cabinets should wherever possible be dispersed with protection provided by continuous flow dry air pressurisation. *Secondary or distribution cables* from cabinets to distribution points (and smaller main cables) should be jelly-filled, not pressurised. *Aerial cables* should be of similar design but with an in-built self-supporting catenary wire in the figure-of-eight mode.

Many modern telephone switching systems incorporate *concentration stages* which can be either *co-located* (i.e. in the same building as the rest of the exchange) or *remote* (i.e. many kilometres away, fed by PCM systems back to the main part of the exchange). Remote concentrators need accommodation and power supplies, and have to be maintained. As time goes on it seems probable however that digital concentrators will become more and more compact and will soon usually be accommodated in roadside cabinets like the cable cross-connection cabinets now is general use. Or perhaps concentrators may soon be designed round so few integrated circuits that they will be put into *sealed canisters* which can be joined in to the cable network and installed inside ordinary underground joining chambers and manholes.

DATA SERVICES

Communication of *discrete messages* (data) is actually an ancient practice. In ancient times communication over distances greater than the range of human voice was provided by sight or sound of discrete signals. Further more, the first practical communication system to use electricity, *the telegraph*, is inherently digital. As telegraphy evolved from the original, *manually based* systems to *fully automated* systems, they developed into what are commonly referred to as *message-switching networks*.

As the need for modern electronic data communications increased, it was only natural that the telephone network would be used for *data transmission services*. The availability overshadowed numerous technical shortcomings of a network designed *primarily* for voice communication services. The main *drawbacks* of a conventional telephone network for data transmission are :

1. Need for signal transducers (modems) on analog access lines
2. Limited data rates
3. High error rates (in the older analog network)
4. Inefficient circuit utilisations

As data communications requirements increased, so did the justification for more *cost-effective* data communication solutions. The *first approach* to reducing data transmission costs was to improve circuit utilisation through the use of *packet-switching networks*. The technology for packet switching was pioneered by the Advanced Research Projects Agency of the US government. This agency developed a network referred to as the *ARPANET*, which eventually evolved into what is now known as the *INTERNET*.

A *second approach* to improving data communications involves developing *separate networks* specifically designed for digital transmission (no analog circuits with modems). Because the data communications market place of the time could not support a separate, dedicated network for data, it ran into financial difficulty.

The *third approach* was the Dataphone Digital Service (DDS) offered by AT&T. This service utilises digital technology of the telephone network for strictly data application. *DDS* circuits are dedicated to data services, but the facilities and routes are *shared* with telephone network facilities. A major hurdle for DDS is achieving *digital access* to DDS circuits. If a subscriber is outside the range of digital transmission facilities of a DDS serving office, a voiceband modem over an analog line is required.

The *fourth approach* to satisfying data communications services involved developing means for *directly accessing* the digital transmission and switching equipment of the telephone network. The first widespread approach of the telephone companies for providing universal digital access is the Integrated Services Digital Network. *ISDN provides digital access to digital facilities of the telephone network for voice or data services on a call-by-call basis*.

MESSAGE SWITCHING

As one telegraph system after another was installed in the countries around the world, nationwise *communications networks* evolved. A message could be sent from one point to another even if the two points were not serviced by a common telegraph line. In this case, telegraph operators at intermediate points would *receive* a message on one line and *retransmit* it on another when a telegraph office had several lines emanating from it, the process of *transferring* a message from one line to another was, in essence, a *switching function*.

One of the world's largest message switches was *completely automated* in 1963 when Collins Radio Company installed a computer-based message switch for airline companies of North America. This system and the more recent successors *store* incoming messages directly into a computer memory (disk file) and *forward* them automatically to the appropriate output line when available. Hence this mode of operation is often referred to as *store-and-forward* message switching.

Included with each message is a *header* containing an *address* and possibly *routing information* so the message processor at each node can determine to which output line to switch the message. The processor in each node maintains message queues for each outgoing link, Fig. 42.7. These queues are normally serviced on a first come, first served basis. However, priority information can sometimes be included in each header to establish different classes, or grades of service, thereby allowing *time-critical messages* to be placed at the head of a queue.

MESSAGE SWITCHING AND CIRCUIT SWITCHING

A message switching network is fundamentally different from a circuit switching network in that *the source and destination do not react in real time*. In fact, most message switching networks could deliver a message on a *delayed basis* if a destination node is busy or otherwise unable to accept traffic. *In a message switching network there is no need to determine the status of the destination node before sending a message, as there is in circuit switching*.

Message switching networks are fundamentally different from circuit switching networks in their response to traffic overloads. A circuit switching network blocks or rejects excess traffic while *a message switching network normally accepts all traffic but provides longer delivery times as a result of increased queue lengths*.

Yet another important distinction of a message switching network is that the transmission links are never idle while traffic is waiting to use them. In a circuit switching network, a circuit may be assigned to a particular connection but not actually carrying traffic. Thus, some of the transmission capacity may be idle while some users are denied service. In contrast *utilisation of the transmission links of a messages witching network is directly related to the actual flow of information*. Arbitrarily high utilisation efficiencies are possible if increased store-and-forward queuing delays are acceptable.

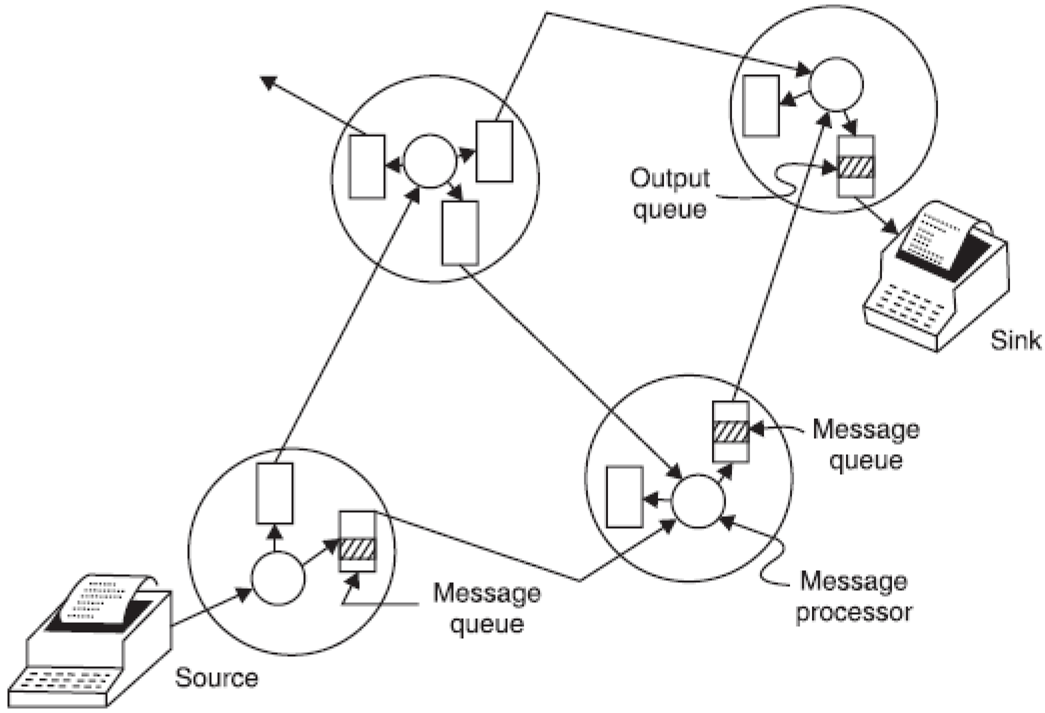


Fig. 42.7 Message switching network

PACKET SWITCHING

Figure 42.8 depicts both the conceptual *structure* and the conceptual *operation* of a packet switched network. A single message at the source is broken up into *packets* for transmission through the network. Included in each packet is a *header* containing address and other control information. Each packet is relayed through the network in a *store-and-forward* fashion similar to a message switching network. At the destination node, the packets are *reassembled* into the original contiguous message and delivered.

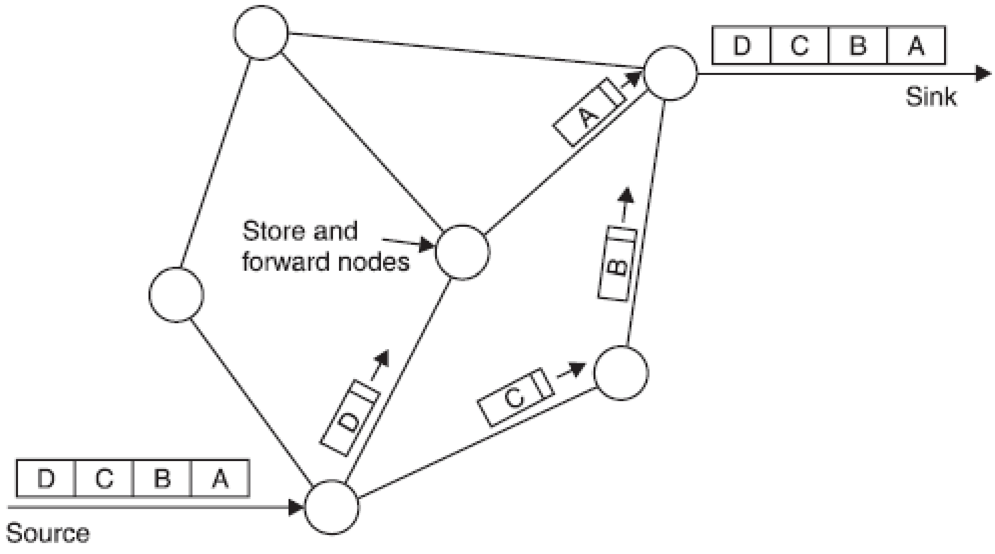


Fig. 42.8 Packet-switching network

The main feature of a packet switching operation is the manner, in which the transmission links are shared on and *as-needed basis*. Each packet is transmitted as soon as the appropriate link is available, but *no transmission facilities are held by a source when it has nothing to send*. In this manner, a large number of relatively inactive sources can *share* the transmission links. In essence, link utilisation is improved at the expense of storage and control complexity in the nodes. Considering the declining cost of digital memory and processing, the increased control complexity becomes less and less significant as digital technology advances.

One particular variation of packet switching *Asynchronous Transfer Mode* (ATM) is designed to specifically support hardware implementations of control intensive functions, thereby supporting very high traffic volumes with low delay.

As the traffic load in a packet switched network increases, the average transmission delay increases correspondingly. In contrast, a circuit switched network either grants service or rejects it. Conversely when only a few circuits are in use in a circuit switched network, much network transmission capacity is idle. *When there is a light load on a packet switched network, the active users benefit by shorter than usual delay times*.

Using automatic repeat request (ARQ) error control, packet switching networks traditionally provided essentially *error-free transmission* for each node-to-node transfer. When errors are detected a *retransmission* is requested (message NAK). Hence transmitting nodes must hold all transmitted packets in memory until a *positive response* (message ACK) is returned by the receiving terminal. Further more, *an entire packet is usually received and checked for errors before forwarding it to another node*.

Customers typically access packet networks by way of leased lines or dial-up connections. *Dial-up connections* are used by infrequent users, while *leased lines* are preferred by heavy users to achieve constant availability, higher data rates, and possibly low error rates.

Despite the similarity to message switching operations, a packet switching network is different in two important respects :

1. The store-and-forward delay through a packet switched network is relatively short. Thus *interactive communications* can occur in much the same manner as if a dedicated end-to-end circuit is established.
2. A packet switched network does not provide storage of messages, except in an incidental manner while relaying packets from one node to another. The network is designed to provide *switched communication* between two nodes, both of which are *actively involved* in the communication process. A packet switching network does not normally store a message for later delivery to an inactive or busy terminal.

PACKET FORMAT

As indicated in Fig. 42.9, a packet contains three major fields : the header, the message, and the redundancy check bits. Generally speaking, the control information associated with a particular message or link is included in the *header* of a message packet. Some packets may not contain a *message field* if they are being used strictly for control purposes. Although a variety of techniques for generating *redundancy checks* are possible, the most popular technique uses cyclic redundancy checks (CRCs). Basically, a *CRC* is nothing more than a set of parity bits that cover overlapping fields of message bits. The fields overlap in such a way that small number of errors are always detected and the probability of not detecting the occurrence of 2 large number of errors is 1 in 2^M, where M is the number of bits in the *check code*.

A header typically contains numerous *subfields* in addition to the necessary address field. Additional fields sometimes included are :

1. *An operation code* to designate whether the packet is a message (text) packet or a control packet. In a sense this field is a part of the destination address with the address specifying the control element of a switching node.
2. *A source address* for recovery purposes or identification of packets at a destination node that is capable of simultaneously accepting more than one message.
3. *A sequence number* to reassemble messages at the destination code, detect faults, and facilitate recovery procedures.
4. *A length* code to indicate the length of a packet when less than a standard size pack is transmitted.

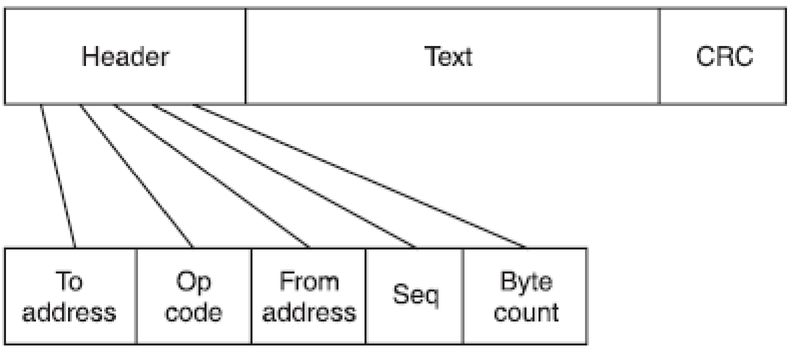


Fig. 42.9 Typical packet format

LAN, MAN, AND WAN

Computers in homes and offices are connected together by *local area networks* (LANs), located within a building or in a campus environment. LANs link computers, printers, scanners and *shared devices* such as modems, video conferencing units and facsimile machines to each other and to the internet (Fig. 42.10). A discrete LAN is typically located on the same floor or within the same department of an organisation.

Each device connected to the LAN can communicate with every other device. The connections between devices may be any of the following : twisted pair, coaxial cable, fibre optics or wireless media. For the most part, devices are connected to a LAN by twisted pair cabling that is similar to but sometimes of a higher quality than that used to tie business telephones together.

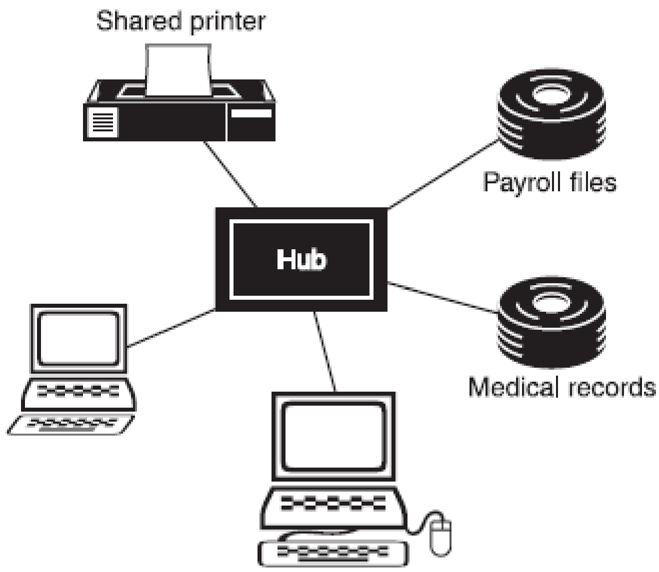


Fig. 42.10 A local area network

We can connect two LANs together that are close to one another (such as within the same building) using a device called a *bridge*. We can connect LANs to each other if they are within a larger geographic region (such as a city) using a *metropolitan area network* (MAN) and we can connect LANs separated by any distance using a *wide area network* (WAN). The basic first step to getting all computers set up so that they have access to all other computers is to get a LAN established for each group of computers. MANs and/or WANs can then be used to connect the LANs together.

A MAN occupies a middle ground between LANs and WANs. MANs cover greater distances at higher data rates than LANs, although there is some *overlap* in geographical coverage.

The primary market for MAN is the customer that has high capacity needs in a metropolitan area. A MAN is intended to provide the required capacity at lower cost and greater efficiency than obtaining an equivalent service from the local telephone company.

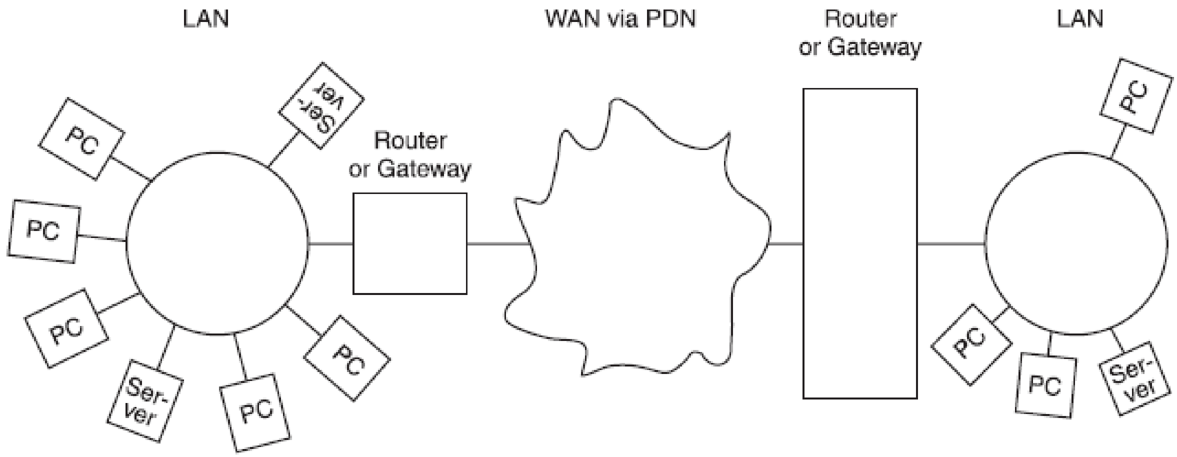


Fig. 42.11 LAN to WAN

WANs cover a large geographic area, may require the crossing of public rights of ways and may rely atleast in parts on circuits provided by a common carrier. Typically, a WAN consists of a number of interconnected switching nodes. A transmission from any one device is routed through these internal nodes to the specified destination device. *High-speed WANs* provide user connections in the 10s and 100s of Mbps, using a transmission technique known as *asynchronous transfer mode* (ATM).

INTEGRATED SERVICES DIGITAL NETWORK (ISDN)

Integrated services digital network is a *single network able to carry and switch a wide variety of telecommunication services, voice, video, data or packets over the public switched telephone network (PSTN)*. It is expected to evolve from an IDN, an *integrated digital network*, which is a telephony network in which digital transmission systems have been fully integrated with switching systems (Fig. 42.12).

ISDN is a service offering that extends access to digital transport facilities and to the signalling network. Access to the *digital transport facilities* occurs on 64 kbps (B) channels while access to *signalling network* occurs on 16 or 64 kbps signalling (D) channels (Fig. 42.13).

Two levels of digital access to the ISDN network have been standardised : basic rate access and primary rate access. As shown, the worldwide *basic rate interface* (BRI) standard is also referred to as 2B+D interface. It consists of two bearer channels for customer voice or data at 64 kbps. In addition, it has one 16 kbps signalling channel. It runs over a single pair of twisted wires between the customer and the telephone company.

The primary interface (PRI) is sometimes referred to 23B+D interface (Fig 42.15). PRI ISDN is a trunk connection. It is installed on the *trunk side* of a PBX or into a multiplexer. BRI ISDN is a *line-side* connection. It connects to the same ports in PBXs as do telephone sets.

Many corporations use PRI ISDN for their direct inward dialing (DID) traffic. The local telephone company sends the caller's names and phone number over the signalling channel. The telephone system captures the information and sends it to the *display equipped ISDN telephone*. Fig 42.15 illustrates a PRI line for transporting *caller ID*. Employees who receive heavy volumes of calls from vendors or who only take calls from certain callers use ISDN to *screen* calls. Calls not taken are forwarded automatically into voice mail.

THE INTERNET

In most cases a LAN or a WAN is not an isolated entity. An organisation may have more than one type of LAN at a given site to satisfy a *spectrum of needs*. An organisation may have multiple LANs of the same type at a given site to accommodate performance or security requirements. And an organisation may have LANs at various sites and need them to be interconnected via WANs for central control of distributed information exchange.

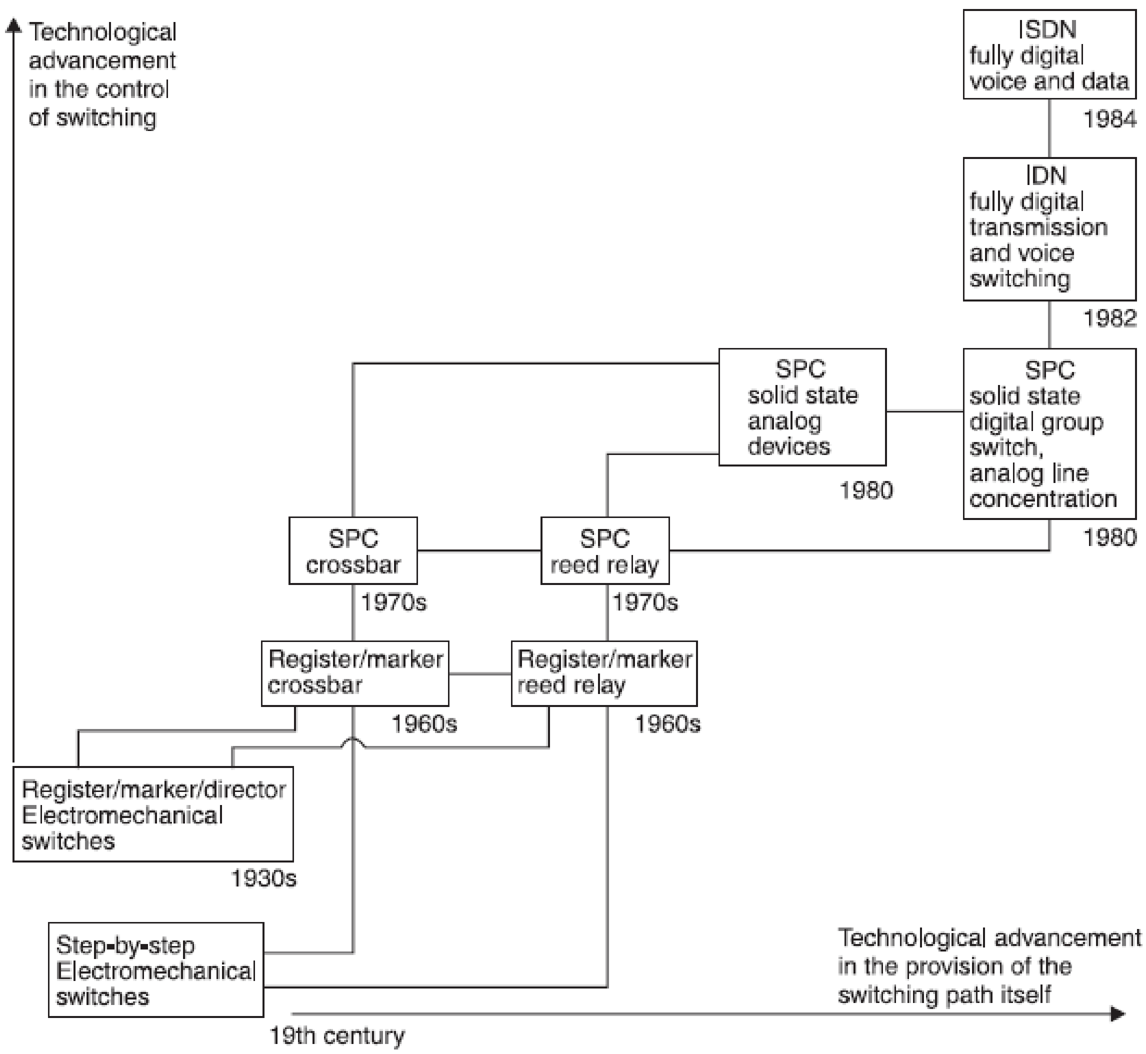


Fig. 42.12 The development of automatic telephony leading to ISDN

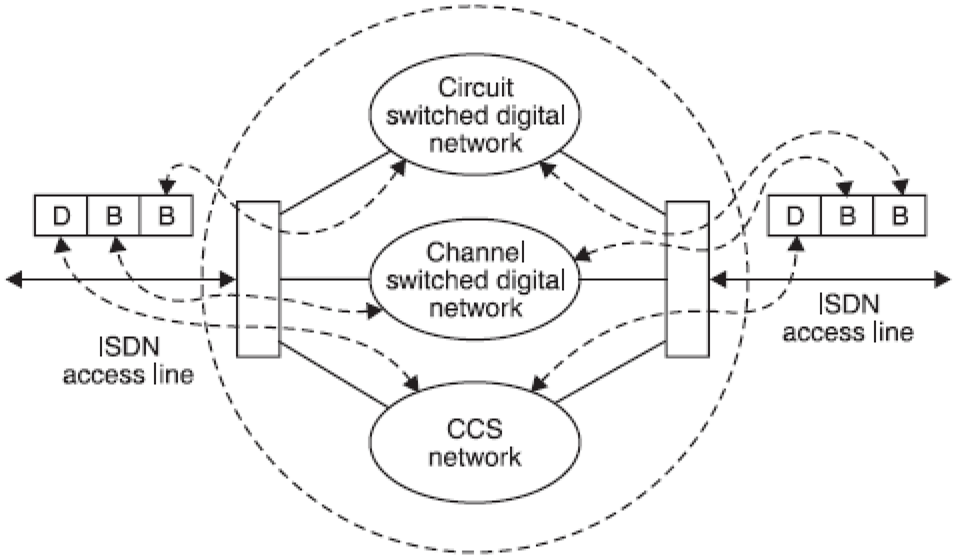


Fig. 42.13 Integrated services digital network access to circuits, channels, leased lines and common-channel signalling.

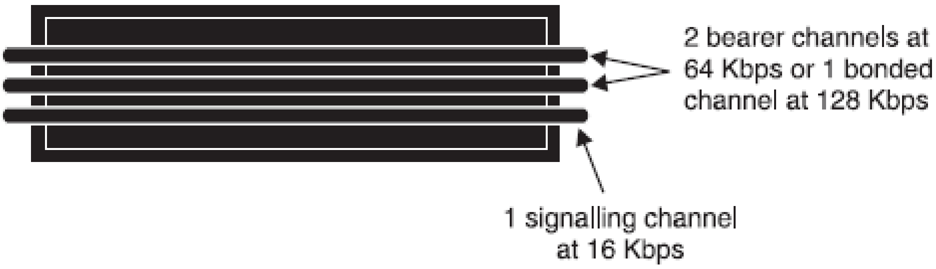


Fig 42.14 Channels and speed of a BRI line

An *interconnected* set of networks from a user's point of view may appear simply as a larger network. However, if each of the networks retains its identity and special mechanisms are needed for communicating across multiple networks, then the entire configuration is often referred to as the *internet*. The most important example of an internet is referred to simply as a *research-oriented packet switching network*. It has served as the basis for the development of internetworking technology and as the model for private internets within organisations. The latter are also referred to as *intranets*.

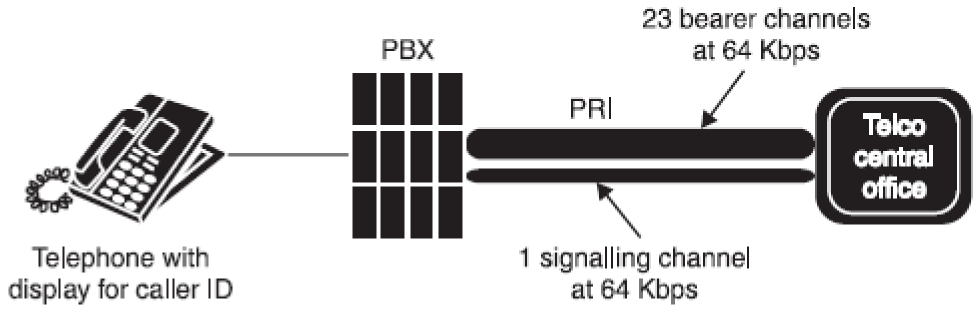


Fig 42.15 A PRI line for carrying caller ID from the telephone company to a PBX

Each constituent *subnetwork* in an internet supports communication among the devices attached to that subnetwork; these devices are referred to as *end systems* (ESs). In addition, subnetworks are connected by devices referred to as *intermediate systems* (ISs). ISs provide a communication path and provide the necessary *relaying and routing functions* so that data can be exchanged between devices attached to different subnetworks in the internet.

Although the internet has grown tremendously, many households still lack PCs and internet access. Thus the area is still ripe for continued growth. It is just as important to have a *PC and internet access* as it was to have a telephone or television 50 years ago. Over time, the telephone and TV came to be seen as necessities. Internet access and a PC in every home *will follow along the same lines*.

The World Wide Web (WWW) organises the resources of the internet, which consist of many computers linked together via high-speed data facilities. These computers contain files that are accessed by *internet surfers*. The collection of files available on the internet are usually accessed by a *web browser*.

EXERCISES

Descriptive Questions

1. What are the advantages of transmitting voice in digital form?
2. What is the major disadvantage of digital transmission?
3. In digital transmission, the bit stream is “reconstructed, retained, retransmitted”. Explain.
4. Differentiate between a time-division and a space-division telephone exchange.
5. Explain the function of a concentration stage.
6. Differentiate between a concentration stage and an expansion stage.
7. What do the BORSCHT functions signify?
8. Briefly explain local distribution networks.
9. Differentiate between circuit switching, message switching and packet switching.
10. With the help of a suitable illustration explain a typical packet format.
11. Differentiate between LANs, MANs and WANs.
12. How ISDNs have evolved?
13. Differentiate between a BRI and a PRI. Give the channels and speed for both.

14. What is the significance of internet and worldwide web?
15. Write short notes on:

    a. Regenerator
    b. Digital trunk exchange
    c. Codec
    d. 2-wire circuit
    e. 4-wire circuit
    f. Primary cables
    g. Secondary cables
    h. Message switching
    i. Header
    j. CRC

Fill in the Blanks

1. With space division exchanges __________________ physical paths are provided, through the exchange, from one subscriber line to another.
2. With digital exchanges switching stages operate by __________________ signals in time.
3. Concentration stages provide the required grade of service with the greatest possible __________________.
4. When a concentration stage serves the called subscriber it is sometimes called an __________________ stage.
5. BORSCHT functions are based on the initials of the __________________ words.
6. To join together a 4-wire circuit and a 2-wire circuit a __________________ coil is used.
7. Any type of time-division switching is a __________________ function.
8. Local distribution networks provide __________________ and as economically as possible.
9. The local network must be __________________.
10. The distribution points are connected by __________________ cables.
11. An entire packet is usually received and checked for __________________ before forwarding it to another node.
12. A header contains numerous __________________ in addition to the address field.
13. LANs that are close together are connected by a __________________.
14. We connect LANs separated by any distance using __________________.
15. A MAN occupies a middle ground between __________________ and WANs.
16. ISDN is a single network able to __________________ a wide variety of telecommunication services.

ANSWERS

Fill in the Blanks

1.	direct
2.	shifting
3.	economy

4. expansion
5. key
6. hybrid
7. one-way
8. service on demand
9. flexible
10. secondary
11. errors
12. subfields
13. bridge
14. WANs
15. LANs
16. carry and switch

CHAPTER 43

MOBILE RADIO SYSTEMS

Of all the tremendous advances in data communications and telecommunications, perhaps the most revolutionary is the development of cellular networks. Cellular technology is the foundation of mobile wireless communications and supports users in locations that are not easily served by wired networks. Cellular technology is the underlying technology for mobile telephones, personal communication systems, wireless internet, wireless web applications, and much more.

Cellular technologies and standards are conveniently grouped into three generations. The first-generation is analog based and still widely used. The dominant technology today is the digital second-generation systems. Finally, third-generation high-speed digital systems have begun to emerge.

WIRELESS LOCAL LOOP

Traditionally, the provision of voice and data communications to the end user over the *local loop*, or subscriber loop, has been provided by *wired systems*. For residential subscribers, twisted pair has been and continues to be the standard means of connection. For business and government subscribers, twisted pair, coaxial cable, and optical fibre are in use.

As subscribers have demanded greater capacity, particularly to support internet use, traditional twisted pair technology has become obsolete. Telecommunications providers have developed a number of technologies to meet the need, including ISDN, and a number of digital subscriber loop technologies, known as xDSL. In addition, cable operators have introduced two way high-speed service using cable modem technology. Thus, *wired technologies are responding to the need for reliable, high-speed access by residential, business, and government subscribers*.

The general appeal for wireless communication is due to its several economic and technical advantages over wireline communication. These include:

1. low installation, operation and maintenance cost;

2. cost independent of distance, thereby rendering it suitable for both high density and short range; application (urban areas), and low density and high range application (rural areas).
3. less effort in maintenance of line;
4. fast deployment of service;
5. higher flexibility and easy expandibility;
6. easy installation over constraint locations like hills, mountains, forests, deserts, sea, rivers etc.,
7. support of mobility i.e. space and time continuity, which means communication at any time and at any location is possible, and
8. wireless communication eliminates the problem of theft of high-resale valued copper.

WLL is not a technology, it is an access method. Variations of existing wireless technologies are used in WLL implementation. Selection of technology will depend upon what application services are required at the user end. The three basic technology options available for WLL are: microwave point-to-multipoint systems, cellular systems and cordless systems.

THE ROLE OF WLL

Fig. 43.1 illustrates a simple WLL configuration. A WLL provider services one or more cells. Each cell includes a *base station antenna*, mounted on top of a tall building or tower. Individual subscribers have a fixed antenna mounted on a building or pole that has an *unobstructed line of sight* to the base station antenna. From the *base station*, there is a link, which may be either wired or wireless, to a *switching* centre. The switching centre is typically a telephone company local office, which provides connections to the local and long-distance telephone networks. An internet service provider (ISP) may be located at the switch or connected to the switch by a high-speed link.

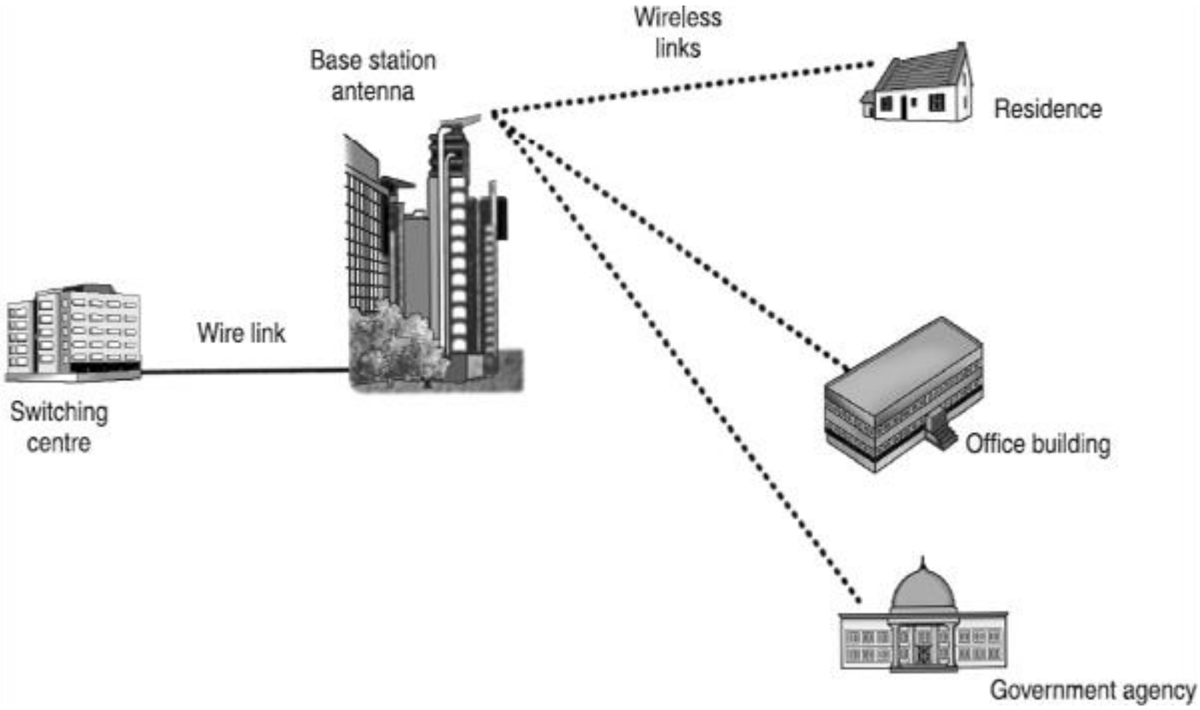


Fig. 43.1 WLL configuration

POINT-TO-POINT, POINT-TO-MULTIPOINT WIRELESS LOCAL LOOPS

Wireless local loops can be point-to-point or point-to-multipoint. In *point-to-point* wireless, each vendor antenna, the base antenna, can support only one customer antenna. With *point-to-multipoint* wireless technology, each base antenna with electronics maintains communications with multipoint customers. Thus one point the vendor's hub site antenna transmitter and receiver-supports multiple customers.

Point-to-point wireless technology was developed by military for applications such as satellites used for surveillance. The technology is now declassified for commercial applications. Point-to-point wireless services are based on bandwidth-on-demand protocols similar to those used in hybrid fibre cable TV systems. In a point-to-multipoint configuration, customers share the bandwidth. It is available *on demand*. The software that manages access to the network is installed on transmitters and receivers connected to antennas. Transmitters are packetized, encrypted, and sorted by software on chips. ATM switches at the hub sites transmit customer data over wireless local networks.

RADIO PAGING SERVICE

Motorola Corporation first introduced *tone-only pagers* in 1956. Tone-only pagers send a tone to a person's pager. People who are paged call a paging operator or an answering service to find out the telephone number and possibly why a person was trying to reach them. Physicians, plumbers and people who needed to be reached in emergencies were early users of tone-only pagers.

Pager sale boomed from the time they were introduced in 1956 until 1998, when growth slowed. When *cellular telephone service* was introduced in the late 1980s, many industry experts thought that this would be the death knell for paging because cellular is a *two-way service*. However, paging sales continued to grow strongly until competition in the cellular industry caused prices for cellular phones to drop. Not only did the prices of cellular phones drop but also paging capabilities were added to the phones. People found that they could carry *one device capable of paging, short messaging and two-way telephoning*. In 1997, for the first time, the number of cellular phones exceeded the number of pagers in service.

Many customers, particularly business people, use their cellular phones fo#r both paging and telephone calls. Personal communication services and traditional digital cellular both support paging and short messaging services (SMS) in the form of brief e-mail messages to users with digital handsets. The gap between functionality in paging and cellular services has narrowed.

Operating Sequence : The radio paging or pocket bell service was put into commercial use in July 1968. The operating sequence of pocket bell service is as follows (refer to Fig. 43.2) :

1. The person who needs to communicate with a subscriber, *dials* the pocket bells' number.
2. The pocket bell system *automatically pages* a subscriber carrying the receiver.

3. The pocket bell receiver *alerts* the subscriber of the page with a tone and flashing LED.
4. The subscriber can *respond* to the page by dialing in from an ordinary telephone number at any time.

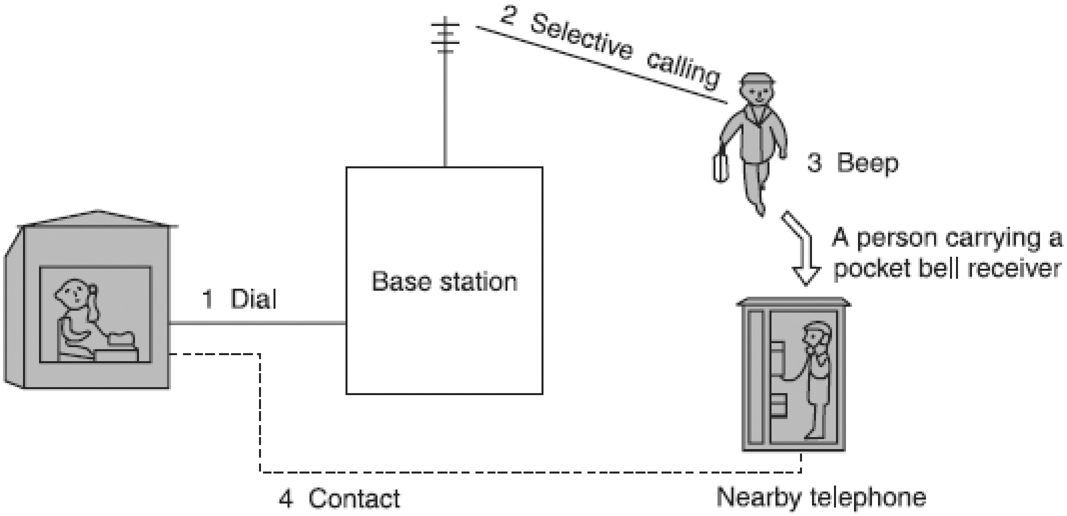


Fig. 43.2 Sequence of "pocket bell" service

Numeric Display Paging Service : To satisfy the demand that a pocket bell can be activated from several offices, a *dual calling service*, which provides different alerting tones to distinguish between calls from two different offices, was put into commercial use in September 1982, and a *numerical display paging service* (Fig. 43.3) was introduced in April 1987. The numeric display paging service uses a new pager that displays a 12-digit message made up of ten numbers (0-9) and four special characters (space, [,], and —) on its LCD display. User can enter messages using touch-tone telephones. For example, the dialing number of a touch-tone telephone used for paging can be entered.

Fig. 43.3 Numeric display paging service

CALL CENTRES

Call centres are the heart of radio paging operation. Call centres receive messages to be paged to the subscribers with the options of automatic call handling or with the assistance of an agent. The call centres are also being used for *customer care services*.

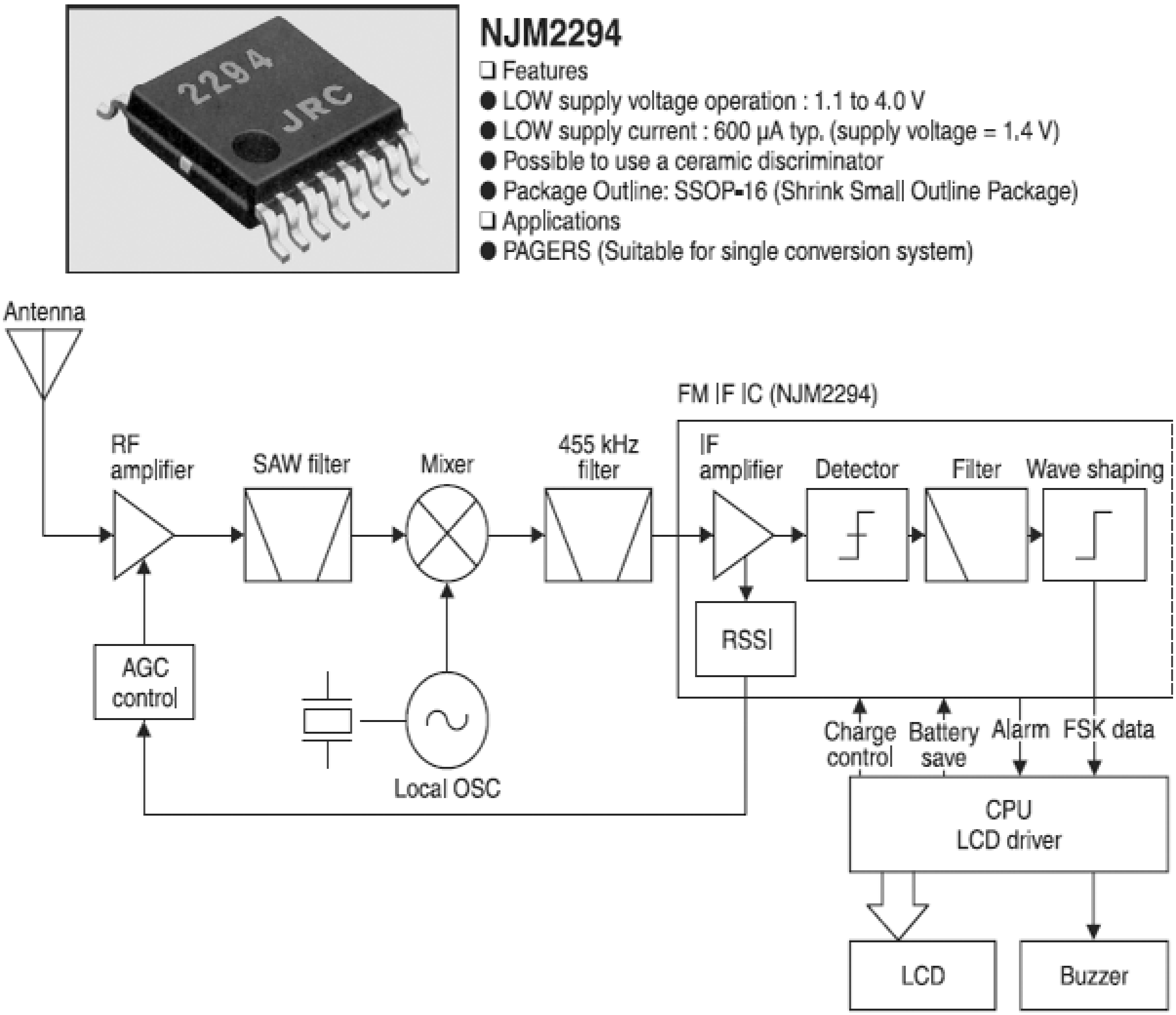


Fig. 43.4 Block diagram of a pager using NJM 2294 IC

Calls are served in two *modes* : operator assisted paging (OAP) and autopaging at different PSTN levels, 9602 and 9601 respectively. *OAP* calls land on the EPABX system with automatic call distribution (ACD) facility. The incoming calls are routed to the operators. The ACD feature ensures that the incoming call is routed to a free or appropriate operator. The operator, on receiving the call, greets the caller with a welcome message. He asks for the pager number and the message to be paged, enters it into the OAP server through a terminal server. Next, he processes the message entered and sends the data to the *paging control terminal* (PCT) for encoding and transmission.

Autopaging calls land directly on the PCT. Caller is welcomed by a voice prompt to leave the message or telephone number. The caller then enters numeric or alphanumeric message using telephones and keypad. The telephone of the caller should be in the DTM mode. The message gets processed and paged to the subscriber automatically.

If the caller has any query, or needs information, calls are routed to the customer care department through a separate ACD group. The customer care department is connected to the main server to access data regarding the subscriber queries and can answer on line. A separate level 9604 has also been introduced exclusively to reach the customer care department.

Radio paging services cannot be provided without a call centre. It is an integral part of the radio paging infrastructure. Apart from paging, call centre is a means to provide quick and quality service to paging subscribers.

VHF/UHF RADIO SYSTEMS

Propagation in the VHF and UHF bands between 30 MHz and 3 GHz takes place in the *tropospheric* mode. The major use of *two-way radio communications* in the VHF and UHF bands is communication between a *fixed base station* and several *mobile units* located on vehicles, ships, or aircraft in the frequency band 30-470 MHz. Typical applications are in control-tower-to-aircraft communication at airports, fire departments, ship control within harbours, police departments, armed-forces field operations, pipeline and transmission line maintenance, highway maintenance, taxicab and delivery vehicle despatch, and personnel paging systems.

Since these systems operate in frequencies above 30 MHz, their range of operation is limited to within the line-of-sight horizon of the base station, or that much further again if a repeater station is used. Large obstacles such as hills or tall buildings in an urban zone create shadows and odd reflection patterns which make complete coverage of the zone from a single base station difficult. For this reason, and to increase the horizon somewhat, it is usual practice to locate the base-station antenna on top of a hill or building to *gain additional height*.

Dispatch systems for automobiles are usually required to cover as much area as possible, and *omni directional vertically polarised antennas* are usually used to accomplish this, both at the base station and in the mobile units. In some applications, the field of operations is strung out in a line over many miles, and for these systems vertically polarised *multi-element Yagi antennas* aimed along the path are frequently used. This provides little coverage *off the sides* but does provide better coverage along the line upto the horizon. The antennas used on the vehicles are nearly always *short ground—whip antennas* as mounted on top of the vehicle. The longer whips used for the 50 MHz VHF band are not as popular.

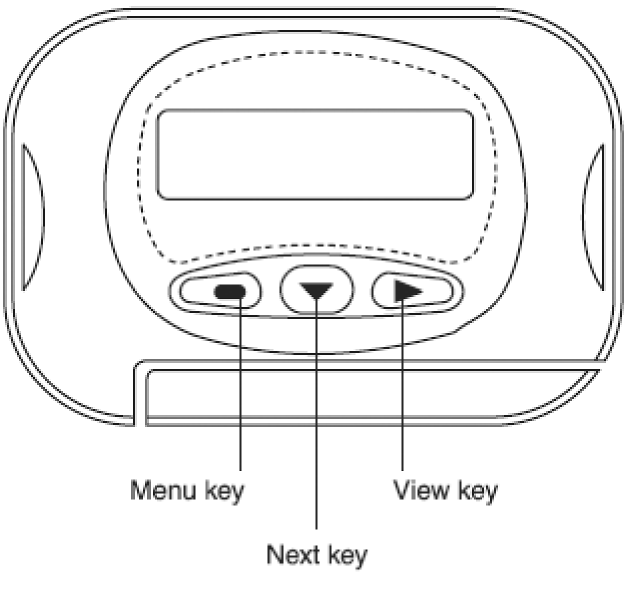


Fig. 43.5 Punwire PK400 alphanumeric pager (3 key. 1 line)

A limited number of channel assignments is available within the spectrum, mostly within the bands 148 to 174 MHz and 450 to 470 MHz. FM operation is preferred, and the *maximum permissible channel spacing* for this service has been progressively reduced from 120 kHz to the 15 kHz presently allowed, so that more channels can be assigned. Because of the narrow bandwidth used, the transmitters and receivers must be very stable, and must maintain their operating frequency within + 5 ppm (parts per million). *Crystal control* is a must if this type of stability is to be realised.

Transmitter power in both the mobile units and the base-station units is usually limited to about 150 W, mainly because of the limited power available from the vehicle system. *Voltage supplies* for mobile equipment range from 12 V nominal for automobiles, 28 V and 48 V for aircraft and 48 V for railway locomotives. The base stations are usually operated directly from the 220 V, 50 Hz power mains, although for some applications *backup battery power* is also provided in case of power failures.

The transceivers are designed to alternately transmit and receive on the same frequency. For aircraft and ship control use, and such systems as police and fire operation, the units may be designed to operate on one of several channels, with *manual switching* between channels provided. Each mobile unit is provided with a *control* head which is usually separate from the main chassis, conveniently located near the operator. The control head provides a power on/off switch, audio volume control, muting threshold control, and handset containing a microphone, a telephone receiver, and a push-to-talk switch. The base station may be self contained and *directly* connected to an antenna near the operator's location, but usually the base station transceiver unit is located with the antenna at a convenient high point and the operator is provided with a *remote control console*. Connection between the base station and the control console is made by means of a pair of wires if the distance is short, or a leased telephone line if the distance is considerable. A simple mobile dispatch system is shown in Fig. 43.6.

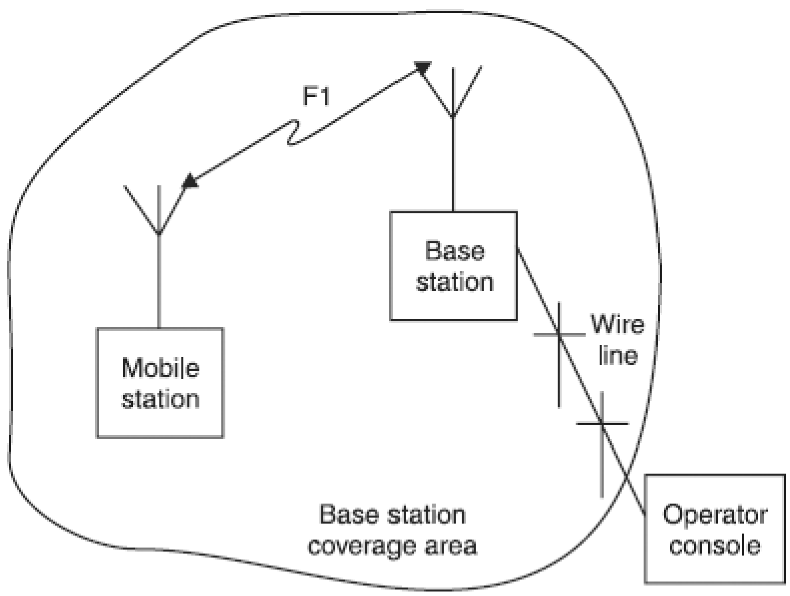


Fig. 43.6 A simple mobile dispatch system

Often it is found that it is impossible to cover the desired area from the single base-station location. In this case one or more *additional base stations* can be established. These may be connected back to the operator's console over wire lines and operated *independently* of each other by the operator. Alternatively, a *radio repeater link* can be established. This requires the use of a *second frequency* for the link between the main base-station and the repeater station. Figure 43.7 shows an arrangement that might be used.

Under normal operation, the base operator can communicate over the local base station on *frequency F1* on any mobile unit within coverage area 1. When it is necessary to reach a vehicle in coverage area 2, he may *turn off* the local-area base station and *turn on* the repeater link on *frequency F2*. Now when the base operator transmits, he transmits on F2 *towards the repeater*. The repeater receives control on F2 and *retransmits* on F1 to the extended-coverage area. When a mobile unit in the extended area transmits, it is received at the repeater on F1 and retransmitted on F2 *towards the base station*.

When a system with several fixed stations is being operated in a repeater mode, the base operator must continuously monitor all the incoming signals from all the repeaters. If a mobile unit should be in an overlap area between two repeater stations, he may key both repeaters, causing interference at the control console. The operator must be able to purposely disable all but one repeater during a conversation. This can be accomplished by sending a *coded tone signal* to turn off the desired repeater stations, and turn them back on when the conversation is complete. The detection characteristics of FM are such that a receiver located in the presence of two co-channel transmissions will suppress the weaker signal. This is known as the *capture effect*.

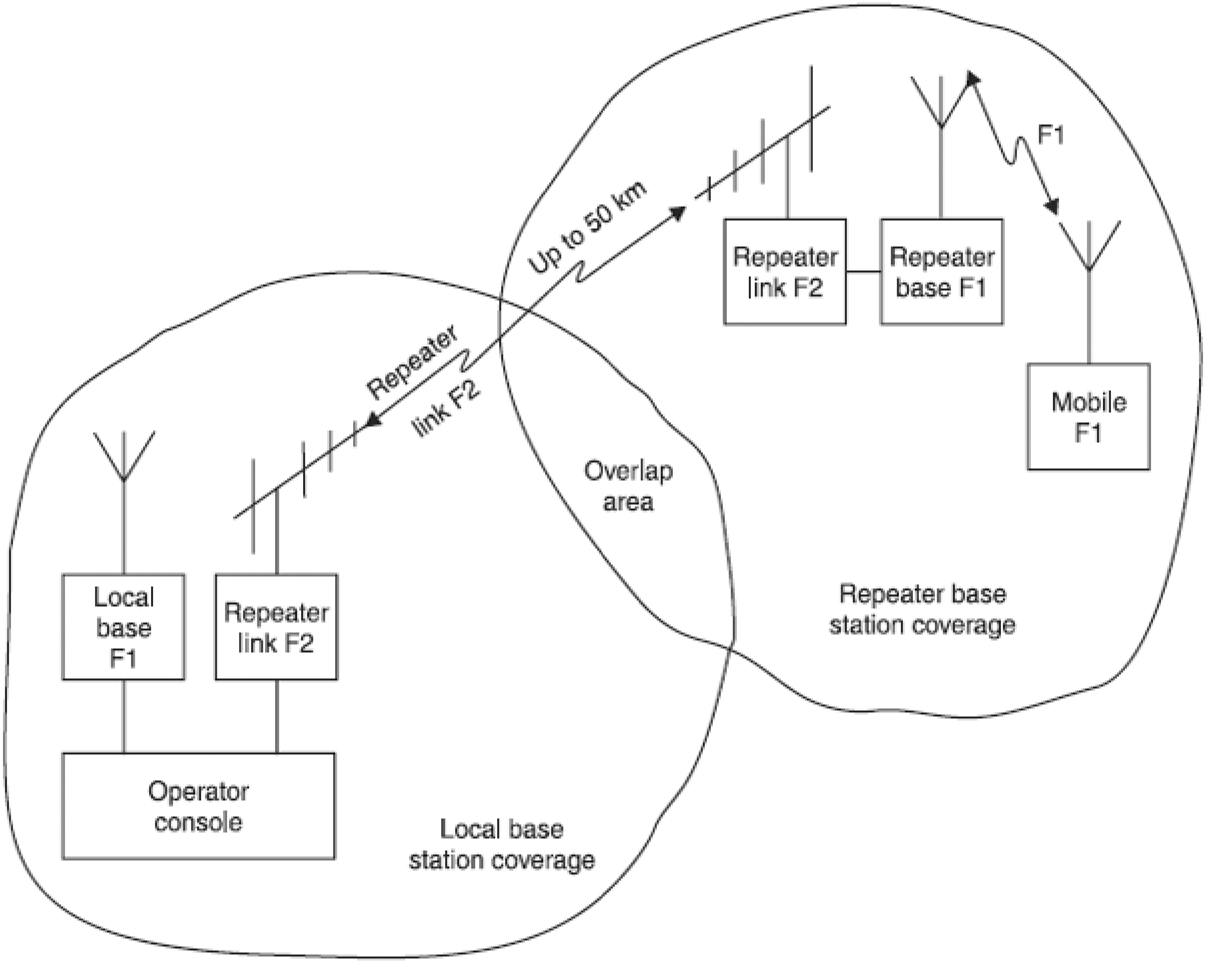


Fig. 43.7 A dispatch station with a repeater station to extend coverage area

LIMITED RANGE CORDLESS PHONES

Many PBX and key system suppliers provide lower priced *home type* or proprietary 900 MHz wireless phones (Fig. 43.8). *The phones only work within range of the antenna in the phone and the phone's base unit*. The 900 MHz phones have a range of about 125 to 150 feet depending on building conditions. There also is an upper limit of about 20 cordless phones supported in the same building.

Proprietary cordless phones have features powered by the phone system to which they are connected. These features, voice mail message lights, multiple call appearances and hold buttons, make the phones easier to use and more functional. *In building wireless phones* are used for the following personnel who often use headsets with their phones:

1. Console operators, to be able to take calls when they step away from their desk for functions such as making copies.
2. Warehouse employees
3. Retail store personnel who can take calls from anywhere in the store.
4. Call centre agents
5. People who work at home

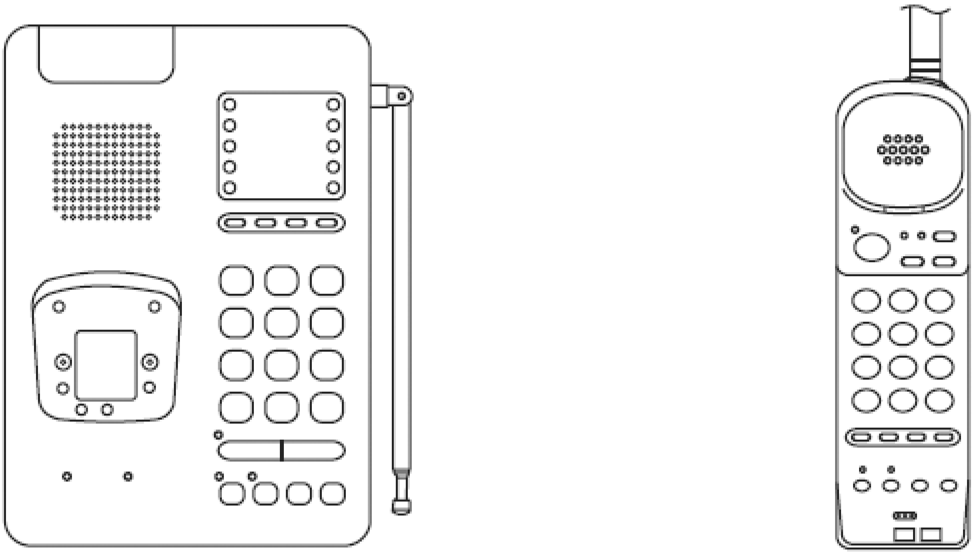


Fig. 43.8 Cordless telephone; base phone and handset

CT1 : The *first generation* of cordless telephone, still very popular, is designed to serve the domestic environment with a *range* of 100 m. The base station transmits on one of eight frequencies between 1.6 MHz and 1.8 MHz, and the handset one of eight frequencies in the 47 MHz band. Frequency modulation is used with a deviation of ±4 kHz at the base station and 2.5 kHz at the handset. *Permitted effective radiated power* at both the base station and the handset is 10 mW.

CT2 : The second generation of cordless telephone. The services, under various names, telepoint, phone zone, etc., were planned to provide for the general public a *lower cost alternative* to the cellular radiotelephone networks which were seen at the time as a businessman's preserve. However, the low market uptake caused their demise in the UK, although services are thriving in some European and far eastern countries.

One hand-portable transmitter/receiver can operate to a local base station installed in the home or office, or through a number of multi-channel base stations with a *range* of approximately 200 m installed in public places. While away from the local base station the subscriber must initiate a call: *calls cannot be made from the PSTN to a CT2 subscriber's handset*.

The operational frequencies are in the band 864.1 to 868.1 MHz and employ time division multiple access (TDMA). Speech is digitised at 32 kbit/s, stored, and then transmitted at 64 kbits/s in 1ms slots. This leaves the alternate 1 ms slots available for the digitised and stored speech of a reply. Duplex operation is achieved in this way on a single radio frequency.

CT3 : The third generation of cordless telephone, *Digital European Cordless Telephone* (DECT) is a pan-European system. DECT operates in the 1880 to 1900 MHz band. It offers data handling facilities and the ability for a subscriber to receive calls while away from the local base station. The techniques are similar to those used for GSM (Global System for Mobile Communications) although, because the mobile is virtually stationary, the constraints on data transmission are less severe and no hand-off is required. The 20 MHz RF bandwidth is divided into 13 carriers spaced at 1.7 MHz intervals, each carrier containing *12 TDMA channels* with GMSK (Gaussian Minimum Phase Shift Keying) modulation.

CT3 is developed from DECT and operates in the band 800 to 900 MHz. Each 8 MHz section of that band is divided into 1 MHz blocks, each containing sixteen 1ms time slots.

CELLULAR COMMUNICATIONS

When an existing system undergoes a unique improvement, it is wise to give that system a *new* name. This is what happened with the *mobile two-way telephone communications system*. Under the original plan, a single medium-power transmitter was placed at the centre of population in an urban community having a *service area* of about 50 miles in diameter. The central transceiver had a *capacity* of between 100 and 500 channels; however, because mobile telephones are full-duplex, the system was *limited* to about 250 conversations at a time. Although the costs were high and subscriber lists were small, the system served the community very well. As the community and the subscriber lists grew, users experienced long waiting periods before they could get into the system. Most of the systems were owned and operated by the local telephone companies, which were capable of supplying mobile-to-mobile as well as mobile-to-home or mobile-to-business *telephone interconnections*.

The geographical shape of the service area was controlled by the radiation pattern of the transmitting antenna, but problems resulted when the community grew or changed shape. For example, a 500 μV field strength from a 10 W transmitter would serve the bulk of the population well, but provide less-than-ideal service to the *fringe areas* of the city. A 500 μV ring from a 40 W transmitter provides good signal to the entire city. The system designer or installer will generally elect to use the high power transmitter *in favour of good signal strength* within the service area, despite the waste of power in the sparsely populated areas.

The changes to this system had to be planned and implemented over a period of time. The *first step* was to use new *system-ready mobile units* in new installations, which offered capabilities far beyond those of the old system. Once the mobile units were in place, the *second step* was to convert the large service area into several smaller service areas, called *cells*. The cells took on the hexagonal shape of a honeycomb. The one high-power transmitter was replaced by six separate 5 W transceivers tied *together* by land line interconnections through the branch exchanges of the telephone companies. A mobile unit in cell C, A would transmit to cell transceiver, then through the telephone system to cell where the message would be *retransmitted* to a mobile unit in cell C's area. These two steps introduced several advantages. *First*, the cell transmitter and mobile transmitter each operate at a lower power. *Second*, a mobile unit in cell A, one in cell C, and a third in cell F could use the same frequencies without channel interference. The *reuse* of frequencies will increase the system's capacity by as many cells as can use the same frequencies. The only *restriction* is that adjacent cells may not use the same frequencies. *Third*, the system is easily expanded because adding a new cell will not affect any of the established cells. *Fourth*, cells provide better management of the total service area.

Fig. 43.9 A typical mobile phone

These are the most obvious *differences*. Each cell's transceiver is connected through the telephone system to a central mobile switching office MSO for better overall system control. The cells are not independent of each other, or of the total system. Suppose a mobile unit in cell A is using channel 24, and another unit in cell C is also using channel 24. So far, no problem since the same frequencies may be used in widely separated cells. Suppose unit A, being mobile, now moves into cell B. *The MSO keeps track of which frequencies are in use in each cell, knowing that adjacent cells should not use the same frequency*. As soon as cell A gives up control of unit A to cell B, the MSO will change the transmit/receive frequencies of either mobile unit depending on the number of channels in use of each cell. *The communications link will thus continue on without co-channel interference. The switchover takes about 250 ms and will generally go unnoticed by either mobile unit operator*.

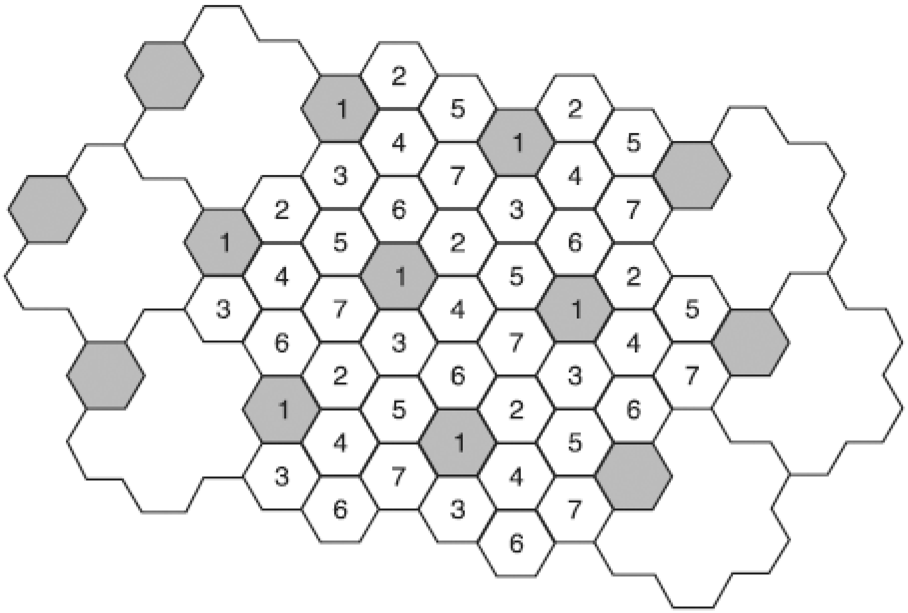


Fig. 43.10 Cellular mobile telephone system. A cellular mobile telephone system divides the area to be served into hexgonal zones or cells. Each cell uses a different set of frequencies to its immediate neighbours. The cells are grouped into blocks in this illustration, blocks of seven cells. The pattern of frequencies is repeated in each block. Thus the tone-filled cells in this illustration use identical frequencies. There could be more cells in a block than illustrated here.

A *step-by-step* trip through a complete transaction will clarify other differences from the original mobile telephone systems. To begin, each cell transceiver sends out an *identification signal* of equal strength for all cells. When a mobile unit operator picks up the handset, a scanning system in the unit measures the signal strengthof all cell's identification code signals. The mobile unit then sets up contact with the cell having the *strongest signal*. The data channel in each cell transceiver, called a *setup channel* operates at a 300 band and uses the ASCII 7-bit code. The mobile unit also sends the cell its *identification code*, which is then passed on to the MSO by telephone line for recognition, frequency assignment, and future billing. The mobile operator then gets a dial tone and initials a call in the usual manner. The call transaction is the same as any standard telephone call.

During the time of the call, the cell transceiver *monitors* all of the active channels in its area as well as in the six surrounding cell areas. It maintains a constant conversation with its six adjacent cells and the MSO. Together, they control which mobile unit is on what channel, which cell is the controlling cell (by comparing the strength of the mobile signal in each area), the location of a mobile unit in each cell, and what action is to be taken when a mobile moves from one cell to another. The *controlling cell* uses a form of reverse AGC; that is, it can order the mobile unit to alter its output so that cell interference is minimised. The mobile output power is maximum at 3 W, but it can be reduced in 1 dB steps (seven steps total) to as little as 0.6 W. All of these *control functions* are carried on without the phone operator's knowledge. The cellular radio network is elaborated in Fig. 43.11.

TRANSMITTING RECEIVING ANTENNA

Consistent with most mobile communications systems the antennas are vertically polarised to guarantee uniform reception and transmission in all directions, regardless of the direction in which the mobile vehicle is travelling. The mobile antenna is a *half-wave vertical whip* usually mounted at the top centre of the rear window of the mobile vehicle and is used for receiving and transmitting.

The *cell transmitting antenna*, again, is a single, driven, vertical, omnidirectional element mounted on top of the transmitting tower (see Fig. 43.12). The *cell receiving antenna system* is shown one-third of the way down the transmitting tower in the figure and consists of six half wave vertical dipole antennas, each with a 90° corner reflector. The position of the dipole antenna with respect to the corner reflector gives each assembly a 60° beam-width radiation pattern ($6 \times 60 = 360$) with about 17 dB gain. The array of the six directional receiving antennas is used to *locate* the mobile unit within the cell area and to *compare* the signal strengths being received by several cell transceivers to determine which cell will maintain *control* of the transmission. In this manner, the cells can detect location of the mobile within a 60° arc and can also determine in which cell the mobile unit is located.

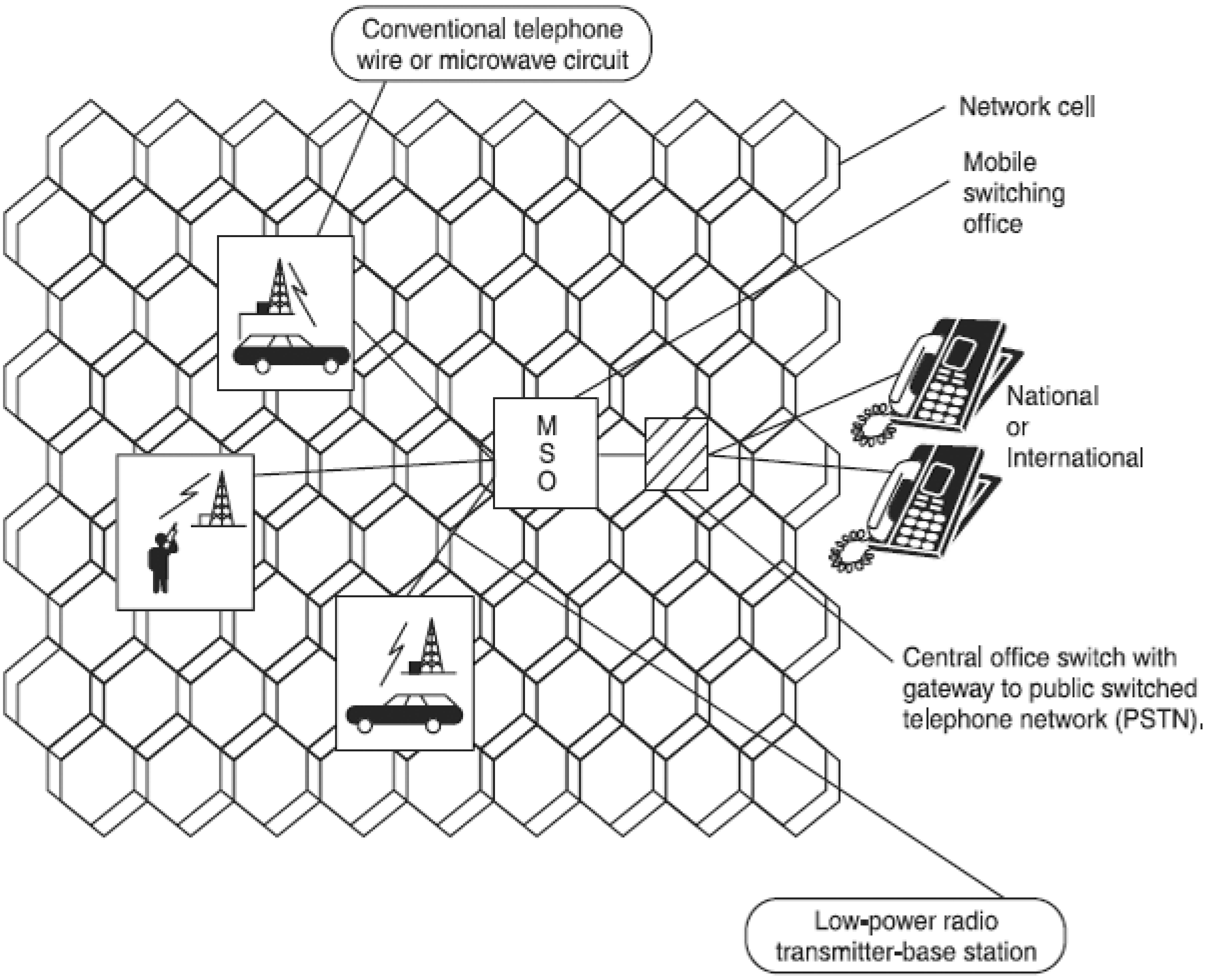


Fig. 43.11 In a cellular radio network each cell contains a base station that is connected to a mobile switching office by a leased terrestrial or microwave circuit. Each base station sends and receives signals from mobile units located in their cell. The MSO handles channel allocation and call switching, and provides a gateway to the public switched telephone network,

thus allowing calls to be switched between the two networks. Because only adjacent cells use different frequencies, a large number of users can be accommodated by a cellular network even though only a limited frequency bandwidth is used.

DIGITAL CELLULAR PHONE BLOCK DIAGRAM

The voice to be transmitted enters the microphone and is input to the vocoder. This coder digitises the audio and then uses complex algorithms to minimise the amount of data needed to represent the information to be sent. This data stream is then input to a channel coder that adds additional bits needed for channel coding. This data stream is then sent to the Pi/4 DQPSK (Differential Quadrature Phase Shift Keying) *modulator* and *up converted* to the cellular frequency band. This RF carrier with modulation is then *filtered* and *pulsed out* at the appropriate time.

On the receiving side a transmitted carrier is *received, filtered, and down-converted* to an intermediate frequency which is *again filtered*. This is then input to an I/Q demodulator and the data recovered. This data still has all of the channel coding information on it so it is then fed into the *channel decoder* to strip off this information. This data is then sent to the *voice decoder* to recover the original speech that was sent. This speech is output through an audio amplifier and then to the speaker.

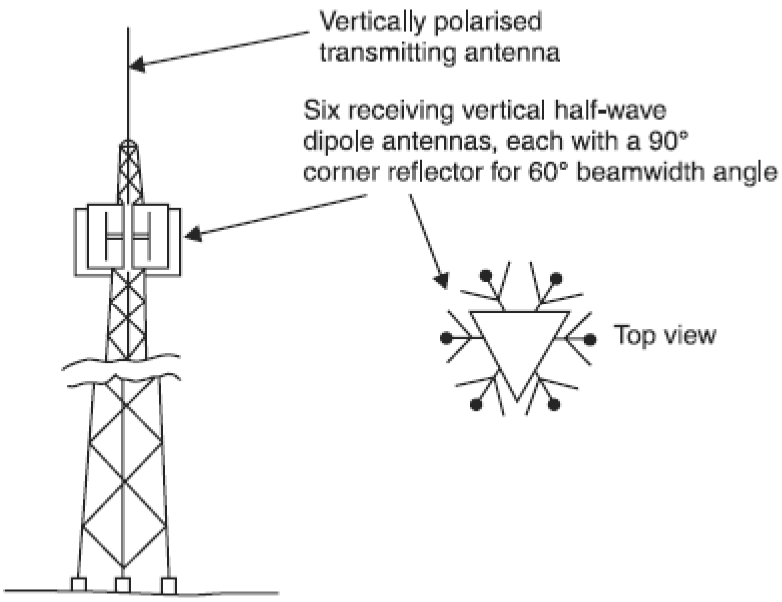


Fig. 43.12 A cell receiving and transmitting antenna tower

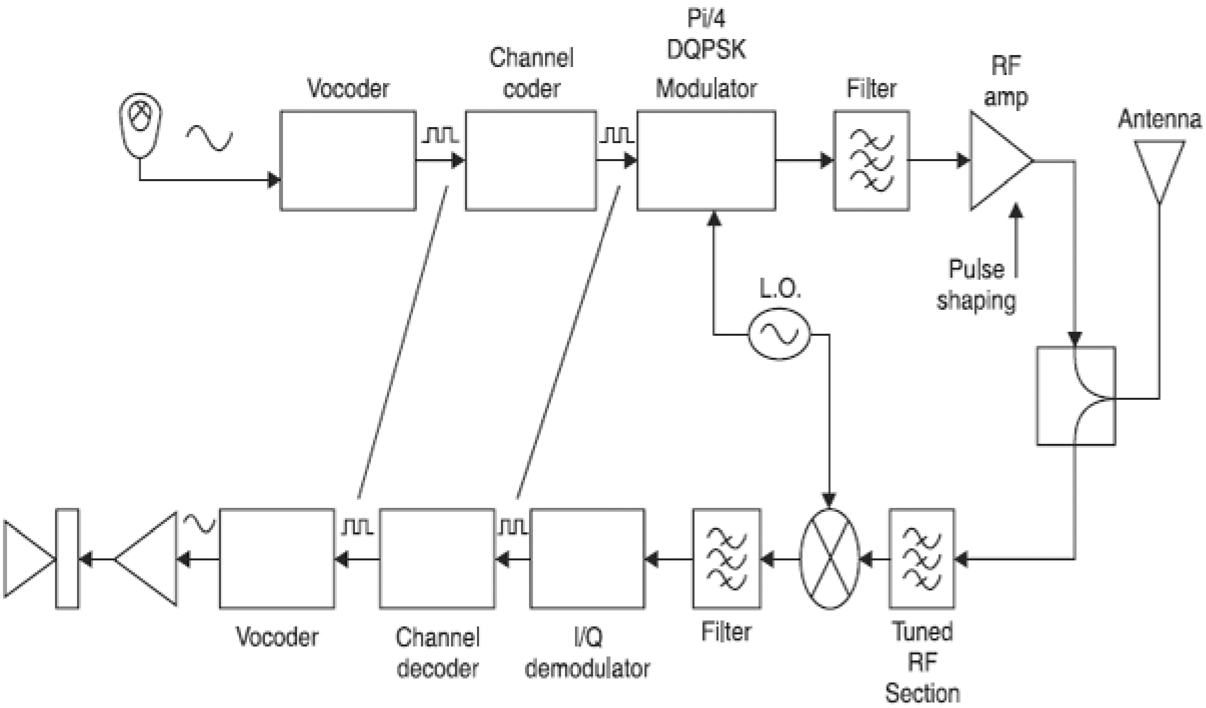


Fig. 43.13 Digital cellular phone block diagram

TYPES OF MOBILE PHONES

There are basically three different types of mobile phones for use with a cellular network: *fixed carphones*, which are permanently attached to the car; *portable phones*, which can be used in the car or outside, and *personal pocket phones*. The difference between the three types of phones concerns their power output and the size of the cell in which they can operate effectively.

Most in-built car phones and many transportables have a *power output* of between 6 and 10 watts and therefore can operate in large cells. The power output *regulates* the strength of the signal the phone can send back to the base station. The farther away the base station, the stronger the signal needs to be; this is also affected by the size of the antenna.

In general, *the best overall reception is obtained from a built-in car phone*. Because of their size, the smaller hand-held phones can have only a relatively small power output, in some cases as low as 0.6 W. These phones are ideal for use in urban areas where the cell size is small and therefore the distance between the base stations is reasonably short.

CELLULAR SYSTEMS

The different cellular systems currently in operation throughout the world are :

Advanced Mobile Phone System (AMPS) : The original system developed by Bell Labs, known as the U.S. cellular standard.

Total Access Communication System (TACS) : Broadly based on the U.S. AMPS system, TACS was chosen as the standard operating network in the United Kingdom.

Nordic Mobile Telephone System (NMT) : The first transnational cellular system to be developed. The NMT cellular network operates across all Scandinavian countries.

Network D and Network C : NetzDor C900 and NetzCor C450 were both developed in West Germany.

Mobile Automatic Telephone System-Europe (MATS-E) : A proposed European System developed jointly by Philips (Holland) and CIT. Alcatel (France).

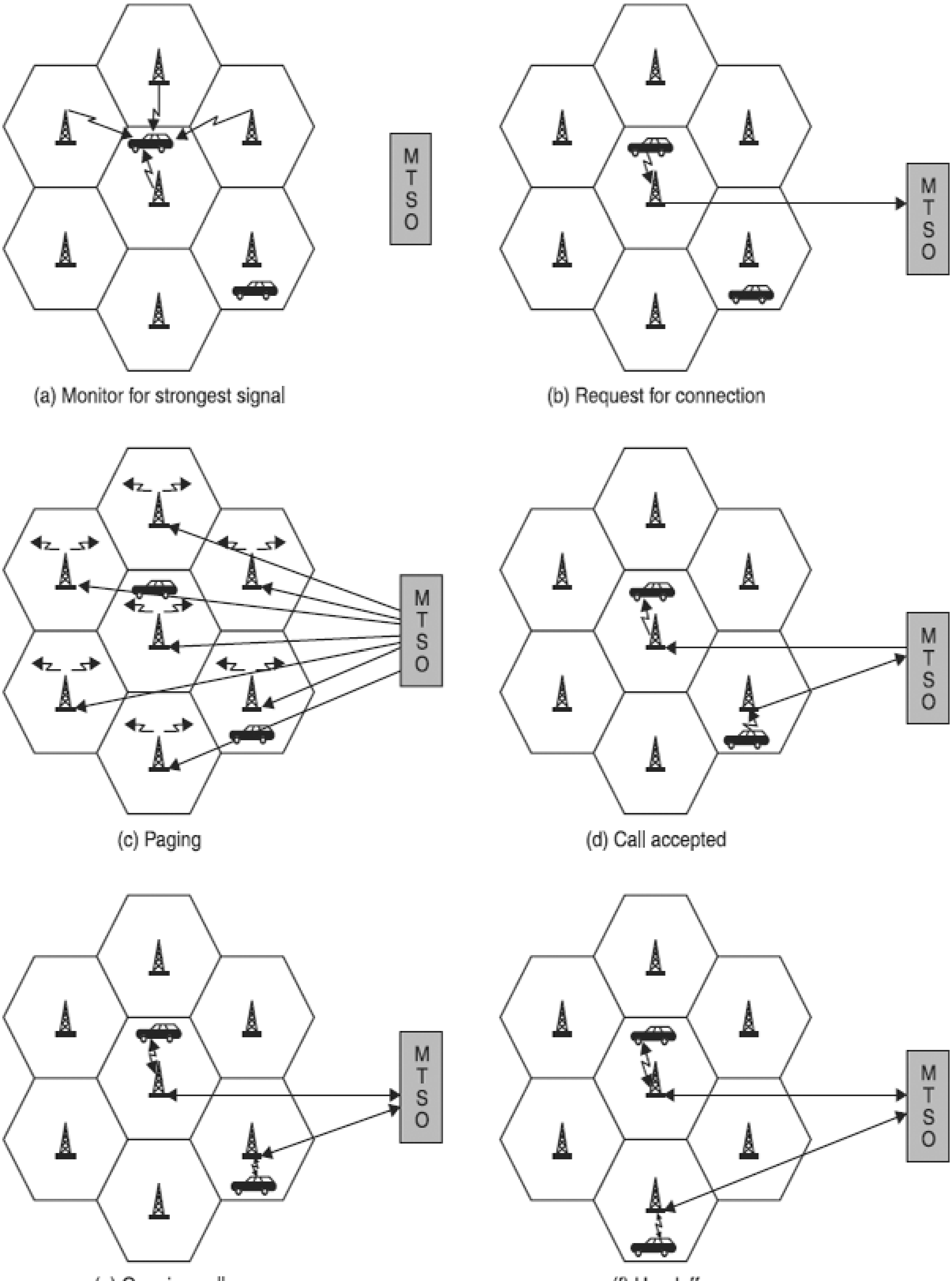
Nippon Automatic Mobile Telephone System (NAMTS) : The Japanese standard for cellular communications.

ESTABLISHING A CALL

Figure 43.14 illustrates the *steps involved* in a typical call between two mobile users within an area controlled by a single MTSO:

1. **Mobile unit initialization :** When a mobile unit is turned on, it *scans and selects the strongest setup control channel* used for this system (Fig. 43.14a). Cells with different frequency bands repetitively broadcast on different setup channels. The receiver selects the strongest setup channel and monitors that channel. The effect of this procedure is that the mobile unit has *automatically selected* the BS antenna of the cell within which it will operate. Then a *handshake takes place* between the mobile unit and the MTSO controlling this cell, through the BS in this cell. The handshake is used to identify the user and register its location. As long as the mobile unit is on, this *scanning procedure is repeated periodically to account for the motion of the unit*. If the unit enters a new cell, then a new BS is selected. In addition, the mobile unit is monitoring for pages.
2. **Mobile originated call :** A mobile unit originates a call by sending the number of the called unit on the preselected setup channel (Fig. 43.14b). The receiver at the mobile unit first checks that the setup channel is idle by examining information in the *forward* (from BS) channel. When an idle is detected, the mobile may transmit on the corresponding *reverse* (to BS) channel. The BS sends the request to the MTSO.
3. **Paging :** The MTSO then attempts to *complete* the connection to the called unit. The MTSO sends a *paging message* to certain BSs depending on the called mobile number (Fig. 43.14c). Each BS *transmits* the paging signal on its own assigned setup channel.
4. **Call accepted :** The called mobile unit recognizes its number on the setup channel being monitored and responds to that BS, which sends the response to the MTSO. The MTSO *sets up* a circuit between the calling and the called BSs. At the same time, the MTSO selects an available traffic channel within each BS's cell and *notifies* each BS, which in turn notifies its mobile unit (Fig. 43.14d). The two mobile units *tune* to their respective assigned channels.
5. **Ongoing call :** While the connection is maintained, the two mobile units *exchange* voice or data signals, going through their respective BSs and the MTSO (Fig. 43.14e).

6. **Hand off :** If a mobile unit moves *out of range* of one cell and *into the range* of another during a connection, the traffic channel has to change to one assigned to the BS in the new cell (Fig. 43.14f). *The system makes this change without interrupting the call or alerting the user*.



(a) Monitor for strongest signal

(b) Request for connection

(c) Paging

(d) Call accepted

(e) Ongoing call

(f) Handoff

Fig. 43.14 Establishing a mobile cellular call

SMART CARD

The smart card or *Subscriber Identification Module* (SIM) card contains a microprocessor and a small amount of memory. With a SIM card, the subscriber can use any PCS *(Personal Communication System)* phone that has a card reader to make a call. The SIM card, the size of a credit card, is inserted into the PCS phone to *activate* the phone. In effect while the card is in the phone the PCS phone becomes *personalised* and becomes the user's personal phone. All the customer's personal data, personal identification number (PIN), services subscribed to, authentication key speed, dialing lists, and so forth are stored in the SIM card.

EXERCISES

Descriptive Questions

1. WLL is not a technology; it is an access method! Explain.
2. Discuss the role of WLL.
3. How radio paging service has evolved? What is its present status?
4. What is the role of call centres?
5. Explain the working of a pager with the help of a block diagram.
6. Describe a simple mobile dispatch system.
7. Describe a dispatch system with a repeater station to extend coverage area.
8. Briefly explain the working of limited range cordless phones.
9. How does cellular communication differ from land line communication?
10. With the help of a suitable sketch, describe a cellular mobile telephone system.
11. What is the role played by transmitting/receiving antenna in a mobile communication system?
12. Draw the block diagram of a digital cellular phone.
13. What are the different cellular systems currently in operation?
14. Illustrate and explain the steps involved in “establishing a call.”
15. Write short notes on :

    a. local loop
    b. operating sequence of a pager
    c. automatic call distribution
    d. tropospheric mode of propagation
    e. mobile switching centre
    f. portable phones
    g. whip antenna
    h. vocoder
    i. hand off
    j. handshake

Fill in the Blanks

1. Call centres are the heart of radio _____________operation.
2. It is usual practice to locate the base-station antenna on top of a hill to gain additional _____________.
3. Voltage supplies for mobile equipment range from _____________.
4. The base operator continuously _____________ all the incoming signals from all the repeaters.
5. The geographical shape of the service area is controlled by the _____________ of the transmitting antenna.
6. Adding a new cell will not affect any of the _____________ cells.
7. The MSO keeps track of which _____________ are in use in each cell.
8. The switchover takes about _____________ and will generally go unnoticed by either mobile unit operator.
9. The best overall reception is obtained from a _____________ carphone.
10. Point-to-point wireless services are based on _____________ protocols.
11. In a point-to-multipoint configuration customers _____________ the bandwidth.
12. Cellular is a _____________ service.
13. The call centres are also being used for _____________ services.
14. Adjacent cells may not use the _____________ frequencies.
15. The _____________ is used to identify the user and register his location.

ANSWERS

Fill in the Blanks

1.	paging
2.	height
3.	12 V to 48 V
4.	monitors
5.	radiation pattern
6.	established
7.	frequencies
8.	250 ms
9.	built-in
10.	bandwidth-on-demand
11.	share
12.	two-way
13.	customer care
14.	same
15.	handshake

CHAPTER 44

FACSIMILE (FAX)

In addition to basic signals consisting of speech, music, or telegraph codes, a telecommunication system is often required to transmit signals of a visual nature. Facsimile means an exact reproduction, and in facsimile transmission an exact reproduction of a document or picture is provided at the receiving end. Television means visually at a distance and a television system is used to reproduce any scene at the receiving end. It differs from facsimile in that the scene may be live (i.e. include movement).

Information is transmitted at a much faster rate in television transmission than it is in facsimile transmission. As a result, television transmission requires a much larger bandwidth, and special wideband circuits are required. The small bandwidth required for facsimile makes it suitable for transmission over normal telephone lines.

FACSIMILE MACHINE

An *input scanner* can be used to transmit images over a telecommunications link to a *remote printer*. This principle has been in use commercially for news photograph transmission since 1930s. Combined *send/receive machines* suitable for office use became widely available in 1960s, and between 1980 and 1987 the number of machines connected to the world-wide telephone network grew from a quarter of a million to two million. Facsimile provides very fast transmission of almost any documentary material without specialist preparation. However, between 2 million and 10 million bits are required for a raster scan of an A4 page; this has to be compared with 20000 bits for a similar-sized page of ASCII-coded characters. This added transmission burden is somewhat alleviated by the ability to tolerate very high error rates.

Fig. 44.1 Personal facsimile machine

The design of facsimile machines has been strongly influenced by the use of the PSTN (public switched telephone network). Firstly, the network was designed for speech, not data. Therefore the power/frequency/time characteristics of the transmitted facsimile signal must be chosen to *suit* the network. Secondly, transmission time is expensive, so effective data compression and channel modulation methods must be used. Thirdly, since the public network is switched, every facsimile machine can in principle be connected to every other. Rigorous development and application of international standards is therefore necessary to ensure that this *potential for interconnection* is not wasted.

Fig. 44.2 shows the block diagram of a typical facsimile machine. When data is read from an input document it is first *compressed* and then *modulated* on to an audio-frequency carrier prior to being *coupled* to the line. The receive path is the *reverse* of this.

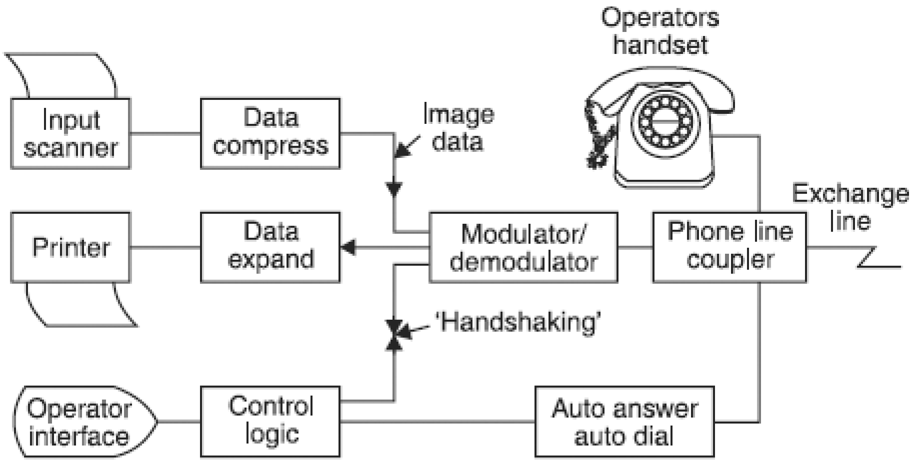


Fig. 44.2 Block diagram of a typical facsimile machine

BASIC FAX MACHINE OPERATIONS

Essentially, a fax machine *scans* original documents, *converts* the scanned images into electrical signals, and *transmits* them over telephone lines to a receiving fax machine. The receiving fax machine in turn *converts* the received signals back into the graphical images of the original document and *prints* them. The basic Group 3 fax machine operations for transmitting a page are presented in Fig. 44.4.

Fig. 44.3 A fax machine is basically a scanner-copier machine that sends and receives graphic images over telephone lines.

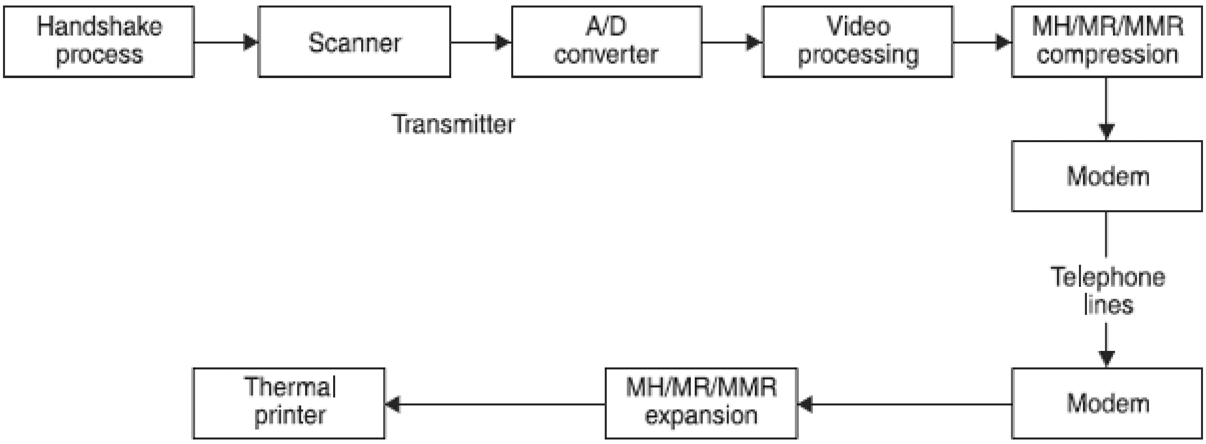


Fig. 44.4 Basic fax machine operations

The handshake process : The sending and receiving *fax-modems* set up the transmission protocols, transmission speed, and other settings between them in a handshake process. If one modem cannot transmit at the highest speed of the other, both modems agree to *fallback* to the next highest speed at which both modems can transmit on the line.

At the transmitting end :

1. Scanning : The images on the page are scanned and transformed into analog signals to begin the transmission process. Either a charge-coupled device or contact image sensor scanner scans the page being sent. A photosensor array of 1728 tiny sensors for A4 paper size (or 2048 for B4) targets very small picture elements *(pixels)* on a line of page, one sensor per pixel, resulting in 1728 (or 2048) bits per line. The array

determines whether each pixel is black or white and accordingly generates a strong or weak electronic signal for that pixel.

A page is scanned line-by-line with all the pixels in a thin strip from 0.13 to 0.25 mm high across the top of the page, or between 10 and 12 scan lines per line of text. Successive strips are scanned until the whole page is converted into a series of electrical pulses. *The amplitude of each pulse represents the brightness of the corresponding pixel*. This scanning operation takes between five and ten seconds per page.

2. A/D Conversion : The scanner signals are converted from analog to digital with typically from one to six bits per pixel. After image processing is complete, one bit per pixel is produced.
3. Video Processing : The processing of the scanner data can be done on the analog scanner signal, the digital data, or both. *It accomodates for the shading, distortions, and other aspects of the original image so that reproduction can be as accurate as possible*.
4. *Shading compensation* checks for non uniformity in the scanner optical system and corrects distortions due to both, light sources as well as nonun iformity in the scanner element.
5. Thresholding : The conversion of the scanner output from grey level to a black-and-white level must also be performed. It may include dithering (or half-toning), a method of generating pseudogrey scales.

   Other video processing techniques include automatic background correction, automatic contrast control, edge enhancement and MTF (modulation transfer function) correction. These can be performed in one or two dimensions. Images may also be reduced or enlarged.
6. Compressing the digital signals : The data compression operation can reduce the picture information by a factor of from 5 to 20, depending on the characteristics of the image. This operation generates code words containing the pixel imformation in compressed digital signals.
7. Modulation : The compressed digital signals are modulated by the modem into analog signals (a tone series) that can be sent over regular telephone lines. *Group 3 fax machines are half duplex and can either send or receive at any time*.
8. Transmission: The analog signals are then transmitted over the phone lines from the sending modem to the receiving modem.

At the receiving end :

1. Demodulation : A modem demodulates or decodes the received analog tone signals *regenerating* the digital signals (bit streams) sent.
2. Decompression : The next step is to expand the digital signals and reconstruct the page's images into black-and-white pixels which represent the pixel's of the page's image.
3. Thermal printing: The thermal printer converts the expanded bit stream into a copy of the original page. The printer's wires are *spaced* 203 to the inch, touching the temperature-sensitive recording paper. For black marking, the wires heat up when high current passes through them. The wires go from *non-marking* (white) to *marking* (black) temperature, and back again in a few milliseconds.

4. Resolution : Standard resolution is 203 lines per inch across and 98 lines per inch down the page. Fine resolution requires twice the number of lines (196 lines per inch) down the page. Most group 3 fax machines include a *high resolution option*.

GROUP 3 FAX MACHINES

Group 3 fax machines now comprise the overwhelming majority of fax machines in operation worldwide. *Group 3 refers to the digital standard that ensures compatibility among fax machines*.

The *Group 1* (1968) analog standard covered four and six minutes per page fax machines, while the *Group 2* (1976) analog standard covered two and three minutes per page machines. The *Group 3* digital standard was first adopted in 1980. It calls for the ability to send an 8.5 × 11 inch page in approximately 30 seconds over a voice grade telephone line. Group 3 fax machines actually do better. They can send an average page of text in 10 to 30 seconds with about 15 seconds for the *initial first page handshake*. The time per page really depends on how many *black markings* (text and graphics images) are present, on their *level of detail* and on the *compression scheme* used.

Due to advancing technology the Group 3 standard has been revised several times since 1980. Most notable advances in VLSI chip technology and DSP have resulted in increased data rates that significantly reduce transmission time.

Group 4 fax machines, which transmit at 64,000 bps, will be suited to computer controlled network communications. First adopted in 1984, Group 4 fax machines are designed for transmission over ISDN. While Group 3 machines excel at stand-alone, person-to-person communications, Group 4 fax machines will be suited to computer-controlled network communications. However the installed base of Group 4 machines today is very small compared to Group 3 machines and they are comparatively expensive.

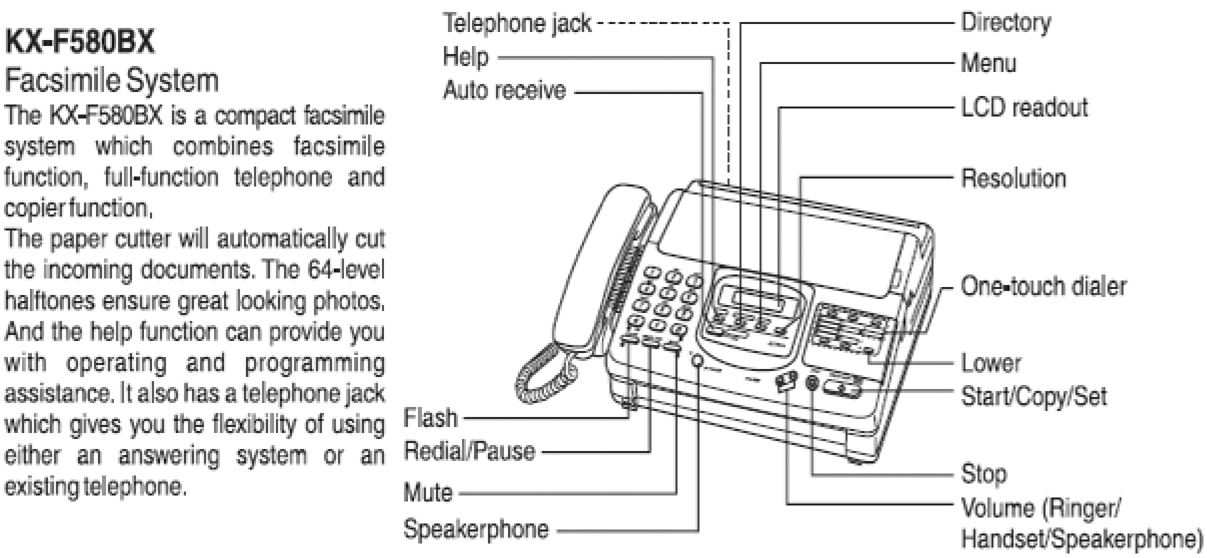


Fig. 44.5 The Panasonic KX-580 BX facsimile system

EXERCISES

Descriptive Questions

1. Draw the block diagram of, and briefly explain, a facsimile machine.
2. Explain basic fax machine operations.
3. Write short notes on :

    a. The handshake process
    b. Decompression
    c. Resolution
    d. Group 3 fax machines

Fill in the Blanks

1. Fax means _____________ reproduction.
2. Fax can be transmitted over normal _____________.
3. Fax provides _____________ transmission of almost any documentary material.
4. If one modem cannot transmit at the highest speed of the other, both modems agree to _____________ to the next highest speed at which both modems can transmit on the line.
5. Shading compensation checks for _____________ in the scanner optical system.
6. Data compression reduces the picture information by a factor of from _____________
7. Group 3 fax machines are _____________ and can either send or receive at any time.
8. Fine resolution requires _____________ the number of lines down the page.
9. Group 3 fax machines include a _____________ option.

ANSWERS

Fill in the Blanks

1.	exact
2.	telephone lines
3.	very fast
4.	fall back
5.	nonuniformities
6.	5 to 20
7.	half duplex
8.	twice
9.	high resolution

CHAPTER 45

XEROGRAPHY

The successful, commercial exploitation of xerography has resulted from the integration of a number of electrostatic process steps. The creation of an electrostatic image, which allows *development by charged, pigmented particles, forms the basis of the xerographic process. These particles can then be transferred to plain paper and subsequently fixed to provide a permanent copy. The generation of the electrostatic image in all xerographic copiers depends upon the phenomenon of photoconductivity, whereby a material's electrical conductivity will increase by many orders of magnitude under the influence of light. By initially charging the photoconductor to a high, uniform surface potential, light reflected from the background areas of the original causes charge decay in the corresponding area of the image. Thus, a charge pattern remains on the surface of the photoconductor corresponding to the original document*.

XEROGRAPHIC PROCESS

The basic steps in the xerographic process are depicted in Fig. 45.1. Charging or *sensitization* of the photoconductor to a uniform surface potential of 800-1000 V is usually accomplised by a corotron. Literally a thin wire, stretched between terminals parallel to the photoconductor surface, and driven at approximately 8000 V, *the corotron emits a corona of ions which deposit on the photoconductor surface. The charged photoconductor must be kept in the dark to prevent discharge*.

Exposure of the sensitized photoconductor to light reflected from the original to be copied generates the required *electrostatic image*. Voltage decay occurs via photon absorption by the photoconductor surface with the creation of an electron-hole pair. These separate under the influence of the electrostatic field, the *electron* neutralising a surface charge, and the *hole*, transported through the photoconductor neutralises the corresponding image charge at the photoconductor–substrate interface. *Selenium obeys the reciprocity law in that it responds to the product of light intensity and time, regardless of their individual values*. Typical light discharge and electrostatic contrast character are shown in Fig. 45.2.

Xerographic development of the latent electrostatic image renders the image pattern visible. To ensure selective development of the image, the black toner particles are charged to a polarity opposite to that of the photoconductor surface. *Toner* consists of finely dispersed carbon black in a thermoplastic polymer matrix. The toner *particles* are small to ensure that reasonable image, edge definition and resolution performance are obtained. Usually transported to the development zone via a carrier, *it is the careful selection of carriertoner pairs that ensures the correct charging and hence development characteristics*.

Transfer of the developed image from the photoconductor surface to plain paper is effected by a *corotron*. Similar in design and process to the charge corotron, positive charge is sprayed onto the back of the paper, which itself is in contact with the developed photoconductor surface. Sufficient fields are generated to ensure that most, but not all, of the toner will transfer to the paper.

Fusing the image into the surface of the paper is accomplished by heat from a radiant fuser, or a combination of heat and pressure from a fuser roll/backup roll combination. It is this step that will dictate the rheological requirements of the thermoplastic resins used in toner manufacture. The process outlined above allows us to visualise the xerographic steps required to produce a copy via a plate photoconductor in the *static mode*.

EXTENSION TO A DYNAMIC COPIER

Dynamic operation of a copier places many *constraints of space and geometry* on the elements of process design and this in turn requires significant demands being made on the xerographic developers and photoconductors in use today. The trend towards faster, less expensive and smaller copiers will ensure that technology development will continue for sometime to come. A schematic representation of a copier is given in Fig. 45.3.

Charging remains as discussed earlier with the rotation of the *photoreceptor* now providing the linear motion underneath the corotron. *Exposure* represents more of a challenge. With stationary platen copiers, the original document is scanned and the reflected light is transmitted via lens, mirrors and exposure slit to the photoreceptor. Some small copiers today incorporate a *moving platen* and strip optics for cost considerations.

Development of the electrostatic image has been extensively studied following the trend towards, faster smaller copiers. Initial copier designs utilised a *cascade development system* which poured developer over the rotating photoreceptor. Increasing process speeds demanded developer with smaller particle size (increasing surface area to mass ratio), to increase the toner carrying capacity. Cascading developer over the photoreceptor in the against mode brought some relief for further speed enhancement but the resultant development system remained bulky. The trend towards *magnetic brush development systems* commenced approximately 20 years ago and immediately offered the advantages of compactness and enhanced developability. The latter, resulting mainly from the *fibrous nature* of the brush and the *density of developer* at the time of photoreceptor contact, has led to almost all modern copiers using this form of development. A more recent development of this concept is the *single-component magnetic brush copiers* used by the Japanese manufacturers.

Apart from the optimization of currents and geometry, the process steps on transfer (and often detack) are direct analogs of the static environment.

Fusing the toner image onto the paper is one of the major contributors to power consumption within the copier and this has led to the abandonment of early, inefficient radiant oven furnace. *Centrally heated polymer-coated steel or aluminium rolls now effect fusing by a combination of heat and pressure*. Release agent fluids are often used on these rollers to prevent toner offsetting from the paper to the polymeric surface.

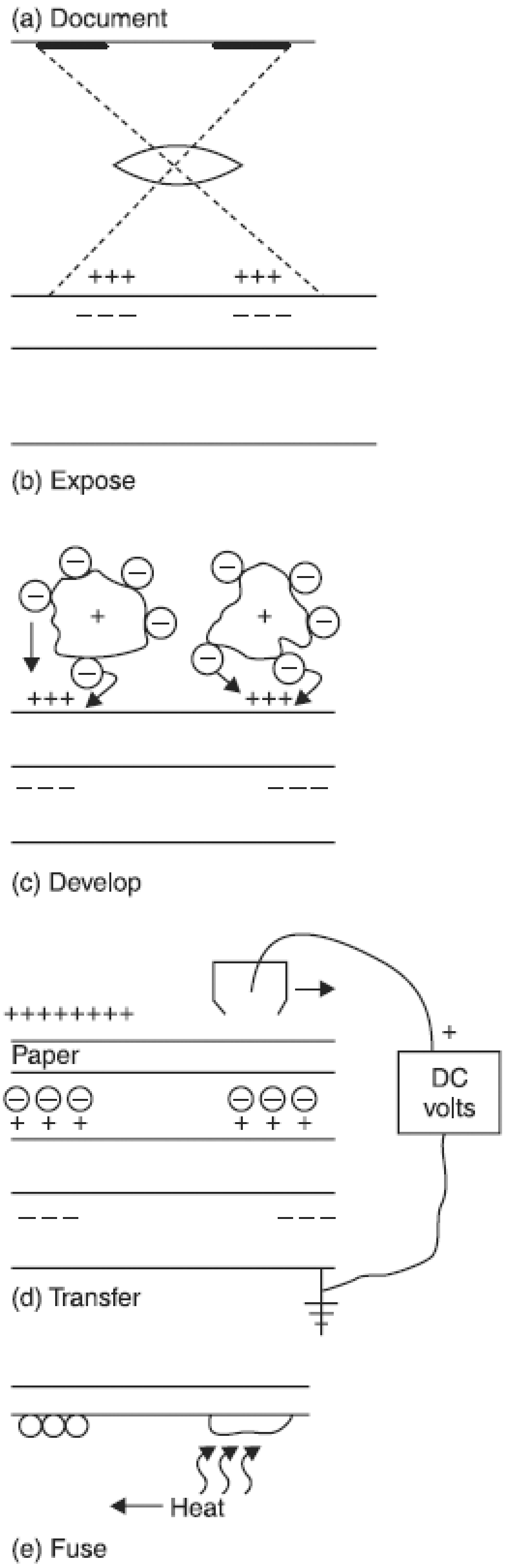


Fig. 45.1 The basic steps of xerography

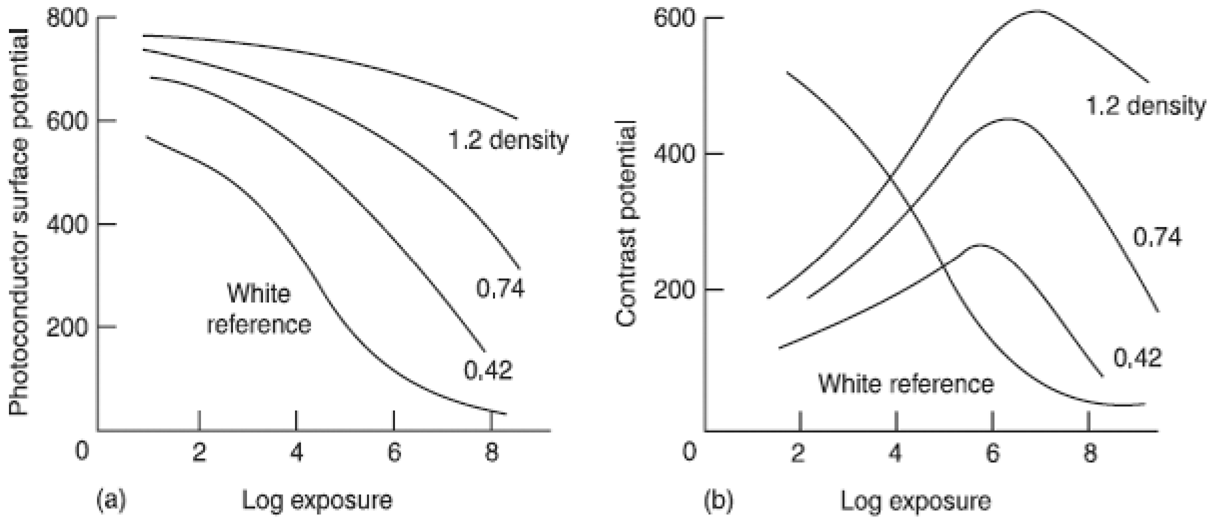


Fig. 45.2 Photoconductor characteristics (a) light discharge (b) electrostatic contrast

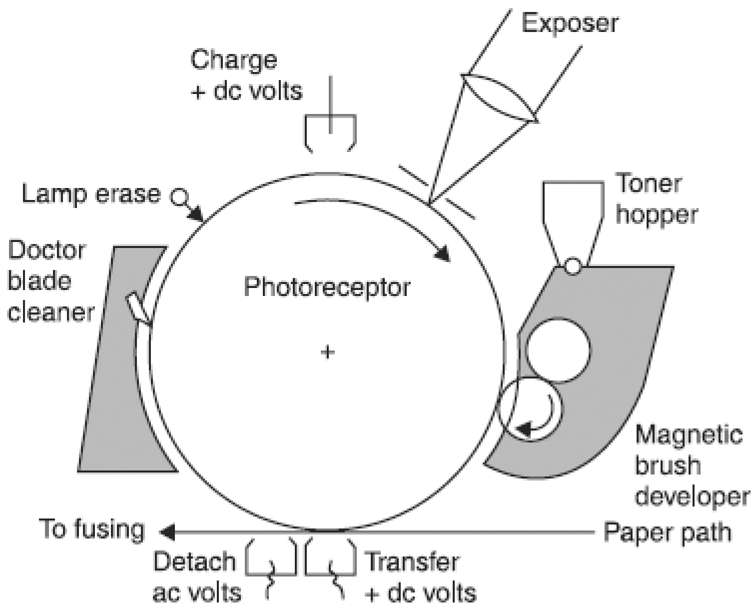


Fig. 45.3 Xerographic copier schematic

In the *dynamic mode* the photoreceptor has to be returned to a virgin condition prior to reaching the charge corotron and hence embarking on another cycle. If this is not the case, *residual voltages* will build up within the transport layer of the photoconductor and *cyclic problems* will rapidly manifest themselves and cause print to print variations. Following the transfer of image toner to a paper, a residual image (some 5.20% of the developed toner) will remain on the photoreceptor surface (Total transfer is avoided as it would increase print background levels). Removal is effected by rotating brushes in some machines, soft webbed fabric in others, but the popular choice for all small and mid-volume copiers today is the so-called *doctor blade*. A straight-edged polyurethane blade is angled against the photoreceptor and this scrapes the residual toner from the surface. The clean photoreceptor subsequently passes underneath an ac corotron or an erase lamp to remove all vestiges of voltage fluctuations prior to charging once again.

An exploded view of a xerographer is given in Fig. 45.4.

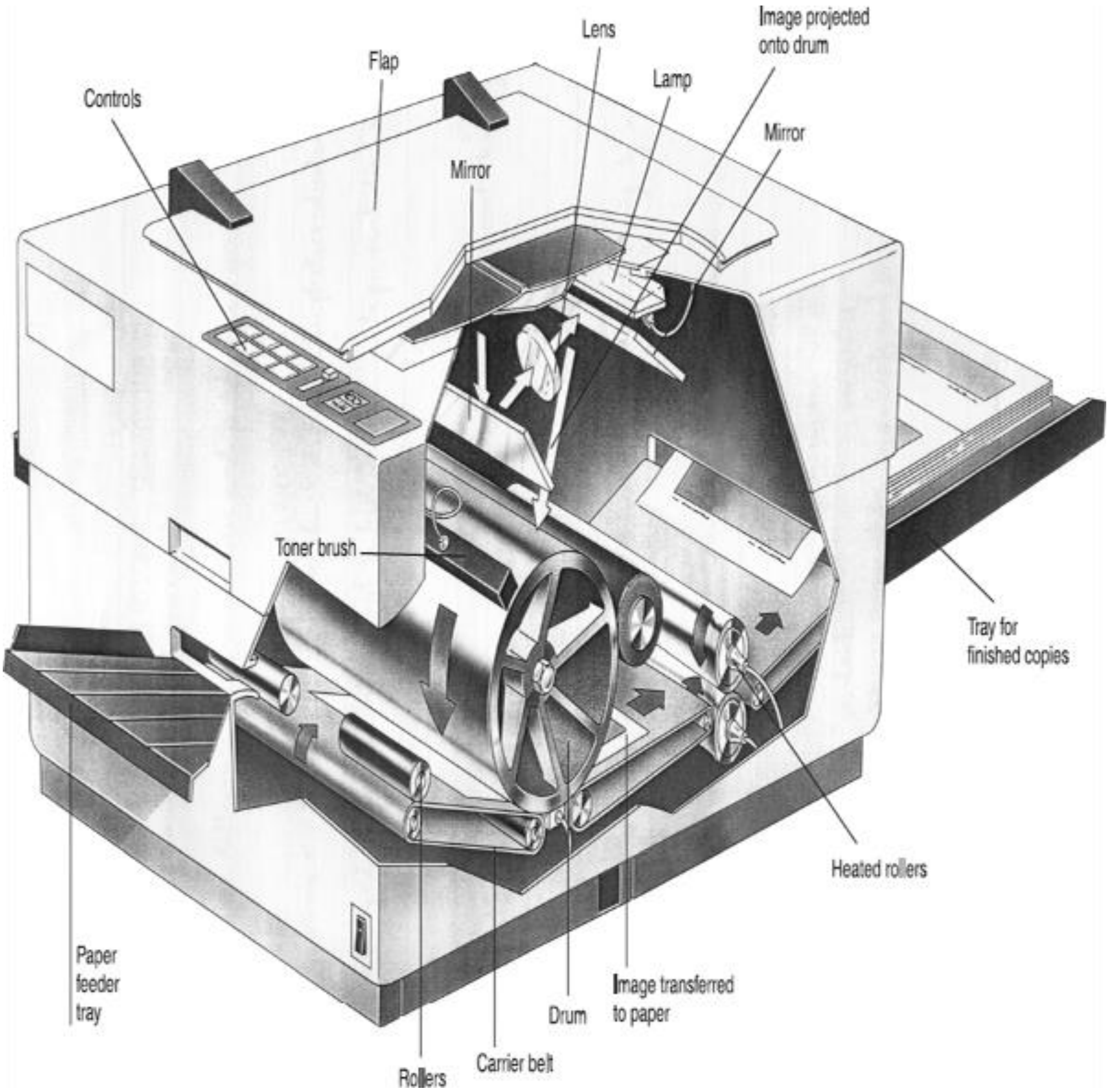


Fig. 45.4 An exploded view of a xerographer

EXERCISES

Descriptive Questions

1. Explain the basic xerographic process.
2. What is the difference between a static and a dynamic copier?

Fill in the Blanks

1. The corotron emits a ____________ of ions which deposit on the photoconductor surface.
2. The charged photoconductor must be kept in the ____________ to prevent discharge.
3. Selenium obeys the ____________ law.
4. It is the careful selection of____________pairs that ensures the correct charging and hence development characteristics.
5. Fusing the toner image onto the paper is one of the major contributors to ____________.
6. Centrally heated polymer-coated steel or aluminium rolls now effect fusing by a combination of ____________.

ANSWERS

Fill in the Blanks

1.	corona
2.	dark
3.	reciprocity
4.	carrier-toner
5.	power consumption
6.	heat and pressure

UNIT V

HOME APPLICATIONS

MICROWAVE OVENS

Every era has contributed its own ideas and conveniences to the kitchen—from the open stove to the closed one, from wood to oil, coal, electricity and gas. And now the age of microwaves is here. Furthermore, microwave ovens were considered a luxury. For a long time, people considered that microwave ovens were used for reheating food only. But as the pace of modern life has got faster and busier by the day, microwave cooking has proved extremely helpful in the efficient preparation of daily food, in the shortest possible time.

MICROWAVES

A microwave is a signal that has a wavelength of one foot (30.5 cm) or less. This converts to a frequency of 984 MHz, so all frequencies above 1000 MHz (1 GHz) are considered microwaves. The frequencies immediately below this border are considered ultra-high frequencies. The upper end of the microwave range contains the light frequencies, about 10^{15} Hz. However, because electronic transmission is so closely geared to *half-wavelength devices*, the practical upper limit is about 300 GHz, where one wavelength is about 0.04 inch (0.1 cm). Smaller devices are being made, but their *power handling abilities* are also micro.

Normally microwaves spread outwards as they travel through the atmosphere and disappear without effect. The *microwave oven* uses microwaves of frequency 2.4 GHz (12.5 cm wavelength) to cook food. Microwave ovens have a *magnetron* usually concealed in the roof of the oven, specifically designed to make use of the energy in the microwaves. Electricity applied to the magnetron tube is used to create *microwave energy*.

Microwaves enter the cooking area through *openings* inside the oven. A *turntable or tray* is located at the bottom of the oven. Microwaves cannot pass through the metal walls of the oven, but they can *penetrate* such materials as glass, porcelain and paper, the materials out of which microwave-safe cookware is constructed. Microwaves do not heat cookware, though cooking vessels will eventually get hot from the heat generated by the food.

4 concave reflectors are located on the left and back sides to concentrate the microwave energy on the food. In state-of-art microwave ovens, the *wave reflector system* (WRS) and *dual-wave emission system* (DES), Fig. 50.1, ensure that the food is always uniformly cooked.

Microwaves reflect off the metal components in the oven (such as the interior walls and the fine screen on the oven door). These metal parts prevent the escape of microwave energy. *All microwave activity remains inside the oven.* When the door is opened or the oven is switched off, the production of microwaves stops instantly.

TRANSIT TIME

Every electronic product with two or more terminals will have inter-electrodes capacitance across the terminals and inductance in series with the terminals. *Both effects limit the usefullness of the device at high frequencies.* Values of 2 pF and 0.02 μ are typical interelectrode reactances that erode the high-frequency signals.

(a)

(b)

Fig. 50.1 (a) The wave reflector system (WRS) (b) The dual-wave emission system (DES)

A second limitation is the time it takes for the electron to travel from one electrode to another electrode called the *transit time.* At microwave frequencies the time for one cycle of RF energy is often shorter than the transit time of the device. The measures required to improve one effect are the *opposite* of what is required to improve the other, and are therefore counter productive. For this reason the principle of *using* the transit time (as in magnetrons) instead of fighting it, has become basic to many of today's microwave devices.

The word magnetron is a conjuction of the words *magnet* and *electrons* and identifies one of the major components, a very powerful magnet. The second major component is a cylindrical copper block, drilled and channeled as shown in Fig. 50.2. The centre opening is called the *interaction chamber.* The holes drilled around the outer edge have a diameter equal to one-half wavelength at the operating frequency and are called *resonant chambers.* There will always be an *even* number of resonant chambers, usually not less than 6 and not more than 16.

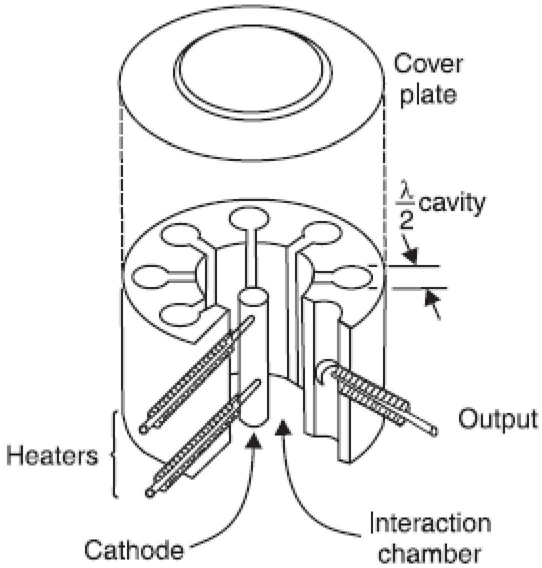


Fig. 50.2 The anode cylinder block of a multi-cavity magnetron

With the magnetron used as a diode the copper block becomes the anode and a directly heated cathode is placed at the centre of the interaction space. The chamber is sealed with top and bottom cover plates and the air is drawn out to form a vacuum. The output connection is a wire loop in one of the chambers that feeds to a coaxial cable fitting on the side wall of the block anode. *Because the anode is exposed to the user it is placed at ground potential and the cathode is at a high negative potential.* The magnetron will only operate as an oscillator (never as an amplifier) and finds its greatest use as a *power oscillator.*

The frequency of the magnetron will remain most stable when any one channel differs in phase from its immediate neighbouring channels by an exact multiple of π/4 radians. Best results are obtained at 4π/4 radians (π radians = 180°). This is called the ? *mode of operation*. To ensure this phase shift of 180° alternate channels are *strapped* together as shown in Fig. 50.3.

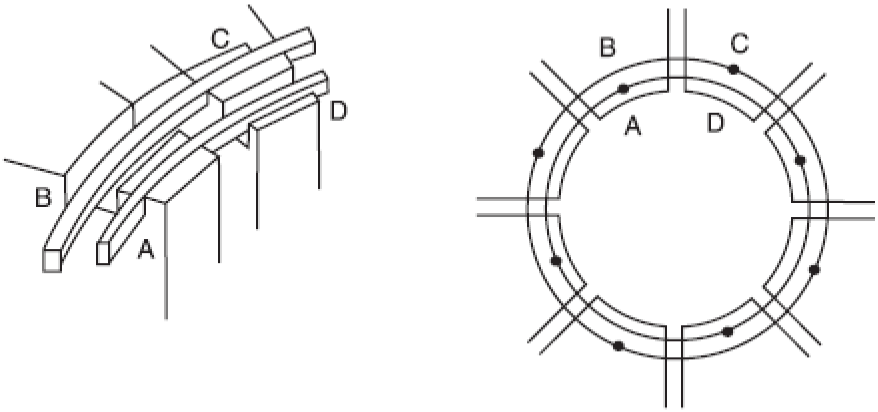


Fig. 50.3 Strapping of alternate anode channel pole pieces to ensure 180° phase shift, ? mode operation. A and C are strapping contacts, B and D are not

WAVE GUIDES

The part of a microwave system that established the theory of operation for all of the other devices is the *interconnecting hardware* called wave guide. The conductors of microwave energy constitute a departure from conventional cables in that *they resemble a coaxial cable with the centre conductor removed*, Fig. 50.4. Microwave energy is carried through the waveguide by *reflection* along its inside walls.

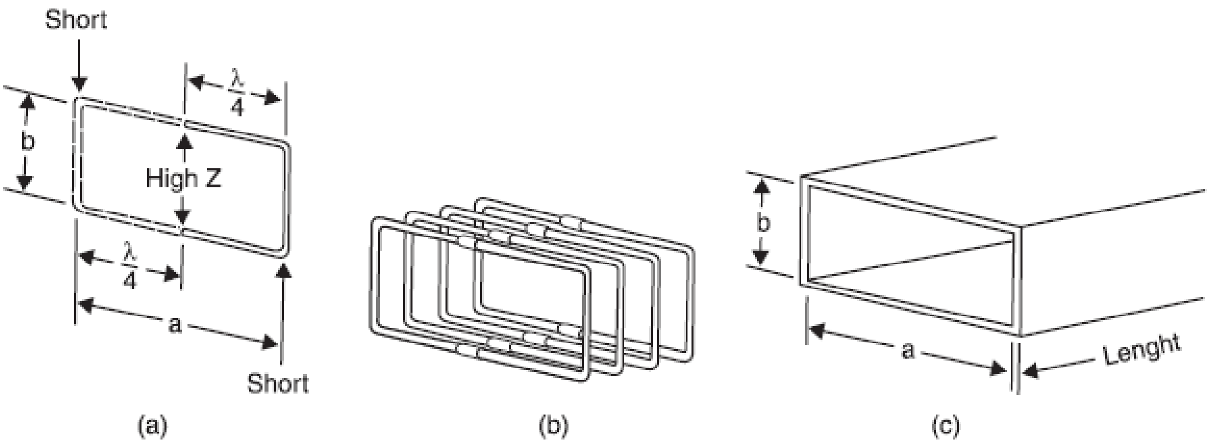


Fig. 50.4 The evolution of waveguides (a) Two quarter-wave shorted sections of transmission line. (b) Multiples of a (c) Standard waveguide with designations (no flange). The narrow side dimension b is 0.4 to 0.5 times the wide side dimension a.

This is possible only if the guide is larger than one-half the length of the applied voltage wave. Thus *the wave guide size is directly related to frequency*. The waveguide factors that deserve attention are :

1. attenuation losses per unit length
2. size selection
3. coupling methods
4. guide impedance and
5. power-handling ability

By transforming microwaves into 3D waves, the system ensures that every inch of food being cooked is *immersed* in 3D microwaves, resulting in food that is cooked more evenly. Microwaves transformed into *three dimensional waves* come out of specially designed waveguides, Fig. 50.5.

Fig. 50.5 Three dimensional microwaves

MICROWAVE OVEN BLOCK DIAGRAM

The block diagram of a microwave oven is given in Fig. 50.6. The mains plug and socket are three-pin earthing type. The fast blow ceramic fuse is of 15 A, 250 V. *Interlock switches are linked with the oven door*. Power will be applied to the mains transformer only when the oven door is closed. At least one interlock switch is in series with the transformer primary, hence even a spot of dirt in the relay or trial, cannot turn the oven on when the door is open.

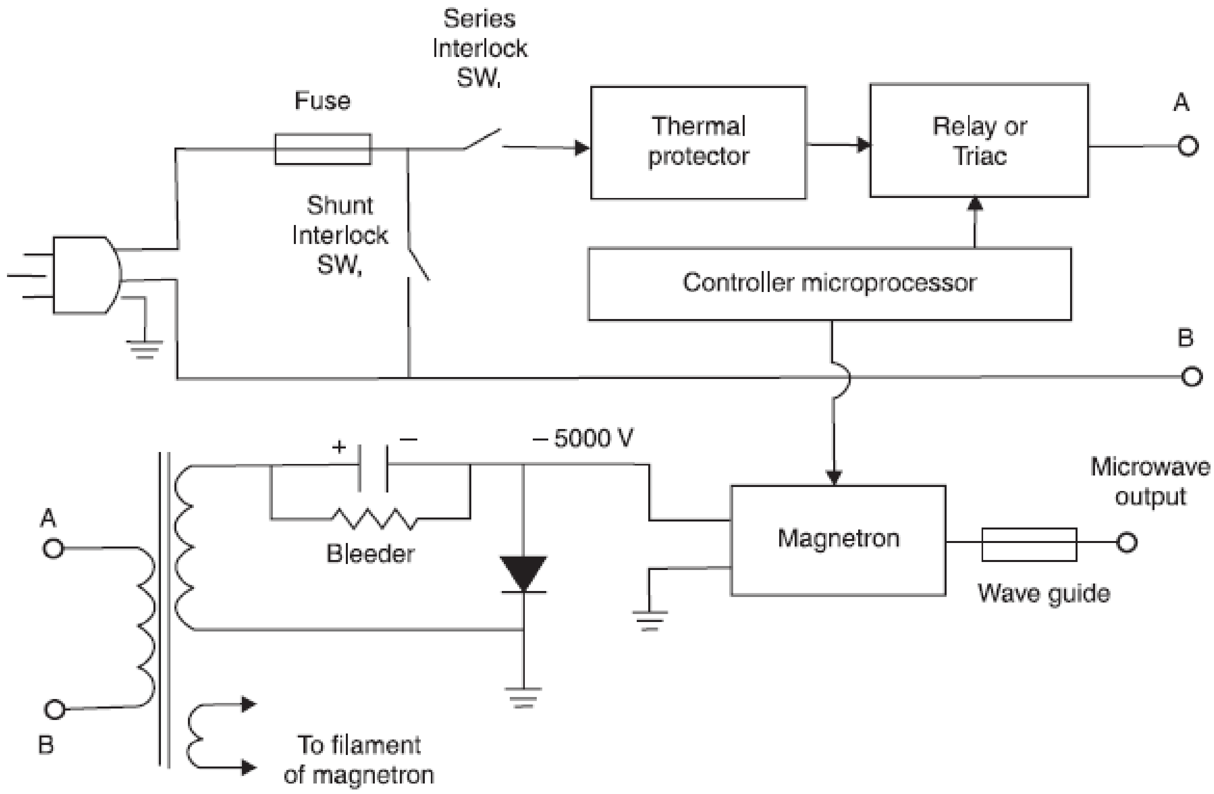


Fig. 50.6 Functional block diagram of a microwave oven

There is yet another interlock across the power supply line. It normally remains open. If the door alignment is not correct it will be activated, putting a short circuit *(crowbar)* across the line and making the fuse to melt. Thus, *the microwave oven is a fail safe device.*

The voltage induced in the secondary winding is about 2000 V (rms) at 250 mA for normal *domestic ovens.* The transformer also has a tertiary winding for the magnetron filament. The *high voltage return circuit* is fastened directly to the chassis through the transformer frame. A *half-wave doubler* configuration is used for the rectifier, with a peak inverse voltage of about 12000 V. One end of the diode is connected to the chassis.

The *bleeder capacitor* (1 μF) should always be discharged before touching anything inside when the cover is removed. The high value *bleeder resistor* is slow to discharge; further it may be open.

The *thermal protector* is a PTC thermistor. The primary current decreases when the temperature rises abnormally. It senses the temperature of the magnetron as it is bolted to the magnetron case and is so connected electrically that its resistance comes in series with the primary circuit.

The *controller* is a microprocessor chip with a clock. It is activated by key-pad switches and sets the cooking time. It senses the temperature and moisture, sets the power levels and runs the display. There are three power levels. For HIGH the microwave generator remains *on* continuously; for MEDIUM it remains *on* for 10 seconds and *off* for 10 seconds;

for LOW it remains *on* for 5 seconds and *off* for 15 seconds. The controller activates the microwave generator using either a relay or a triac.

LCD TIMER WITH ALARM

Most microwave ovens feature at least one timer with an alarm. Older appliances used mechanical timers, but modern microwave ovens and cooking ranges feature electronic timers using digital circuitry. The concept of a timer is sketched in Fig. 50.7. In this system, the keypad is the *input* and both the digital display and alarm buzzer are the *output* devices. The processing and storage of data occur within the digital circuits block in Fig. 50.7 (a).

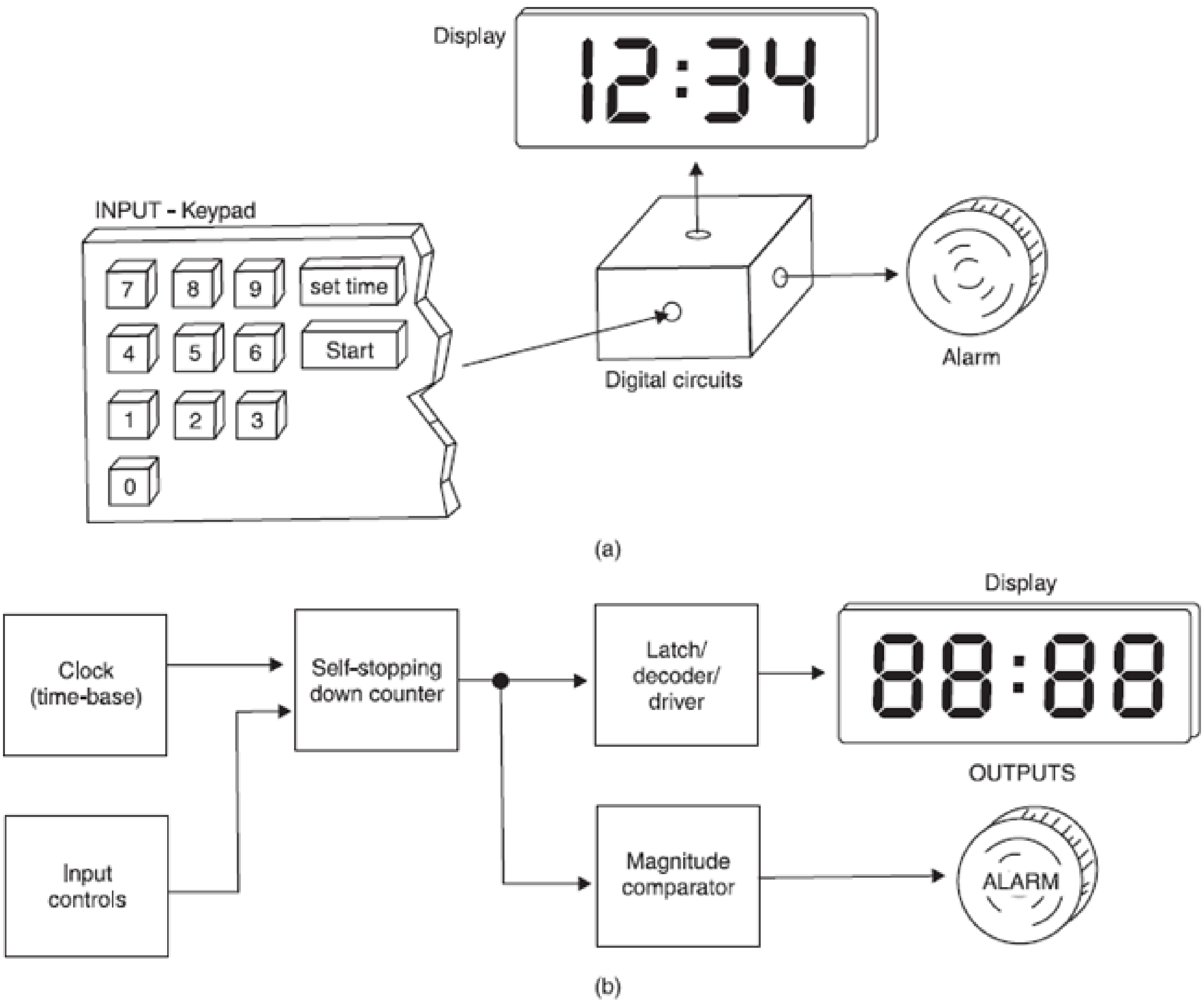


Fig. 50.7 Digital timer system

A somewhat more detailed block diagram of a *digital timer* is shown in Fig. 50.7 (b). The digital circuits block has been subdivided into four blocks. They are the *time-base clock*, *the self stopping down counter*, the *latch/decoder/driver*, and the *magnitude comparator*. The *input controls* block presets the time held in the down counter. The time base is a stable multivibrator which generates a known frequency. In this case, the signal is a 1 Hz square wave.

The *accuracy* of the entire timer depends on the accuracy of the time-base clock. Activating the *start* input control causes the down counter to decrement. Each lower number is latched and decoded by the latch/decoder/driver. This block also drives the *display*.

SINGLE-CHIP CONTROLLERS

Most of us are familiar with general-purpose microcomputers such as the IBM PC and its clones and the Apple Macintosh, which are used in more than half of our homes and in almost all of our businesses. These microcomputers can perform a wide variety of tasks in a wide range of applications depending on the software (programs) they are running. There is a more specialised type of microcomputer call a **microcontroller** which is not a general-purpose computer. Rather, it is designed to be used as a *dedicated* or *embedded controller* which helps monitor and control the operation of a machine, a piece of equipment, or a process. Microcontrollers are microcomputers because they use a microprocessor chip as the CPU, but they are much smaller than general-purpose microcomputers because the input/output devices they normally use are much smaller. In fact, some of the input/output devices—as well as memory— are usually right on the same chip as the microprocessor. These *single-chip microcontrollers* are employed in a wide variety of control applications such as: appliance control, metal-working machines. VCRs, automated teller machines, photocopiers, automobile ignition systems, antilock brakes, medical instrumentation, and much more.

A very typical application of an embedded microprocessor is a microwave oven control system. A block diagram of such a system is shown in Fig. 50.8. If we were to try to analyse all of the machine-language instructions needed to program an actual microwave oven, you would find it overwhelming. Our goal here is to understand how a simple part of a program works and give you a glimpse of what the program does to control the system. In the example of Table 50.1 only a portion of a program is shown in a simplified form that you will find easy to understand. Its purpose is to determine if a non-zero value has been placed in the accumulator. The value in the accumulator represents the number of seconds that the microwave should cook the food. If a non-zero value is in A, it displays the number of seconds on an output port and counts down in 1 second interval until it reaches 0. It then continues with the rest of the program. The program starts executing at address 0000 when power is first applied, which resets the system. The instruction that is generally stored at the reset address is a jump instruction that sends the micro to the main program. The main program in this case starts at 0100, where it makes a decision either to jump immediately to the rest of the program at 010A or to execute the instructions from 0102-0109. In either case, it eventually executes the rest of the program from 010A until it is told to jump back to 0100 and do it all over again.

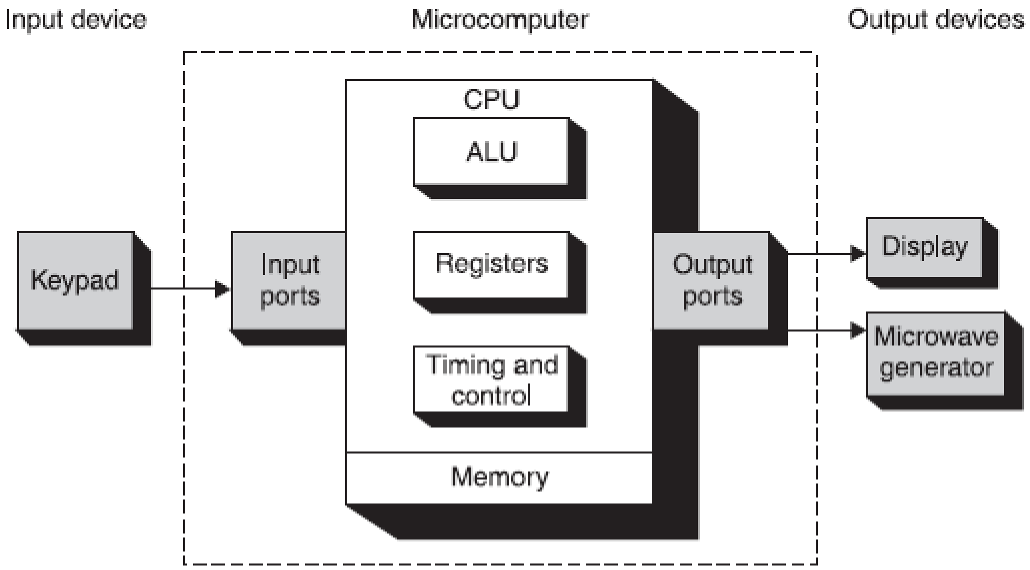


Fig. 50.8 Single-chip microcontroller block diagram

TYPES OF MICROWAVE OVENS

There are two main types of microwave ovens in the market. The first type cooks by microwaves only while the second is a microwave convection oven, which is in fact a combination of ovens. While *microwave ovens* remain popular, there is now a great demand for *combination ovens* also.

Table 50.1 Sample Machine-Language Program

Memory Address (Hex)	Memory Contents (Hex)	Assembly Language	Description
0000	02	LJMP 0100H	JUMP to start of program
0001	01		
0002	00		
0100	60	JZ 010AH	Should we cook the food?
0101	08		
0102	F5	MOV P1, A	Display cook time on port 1
0103	90		
0104	12	LCALL 1_SEC_DELAY	Waste one sec
0105	28		
0106	55		
0107	14	DEC A	Subtract one sec from time
0108	70	JNZ 0102H	Is the food done?
0109	F8		
010A	*****This is where the rest of the program continues.*****		

Most food cooks wonderfully well by the *moist cooking* method in microwave ovens, but certain food requires the *dry heat* produced in normal conventional ovens to turn it crisp and brown. This is when the second type of oven helps. Hence, some of the latest models have a combination of microwave and conventional ovens called the *combo mode of cooking* where the food automatically cooks by means of microwaves and then crisps by the conventional method, Fig. 50.9.

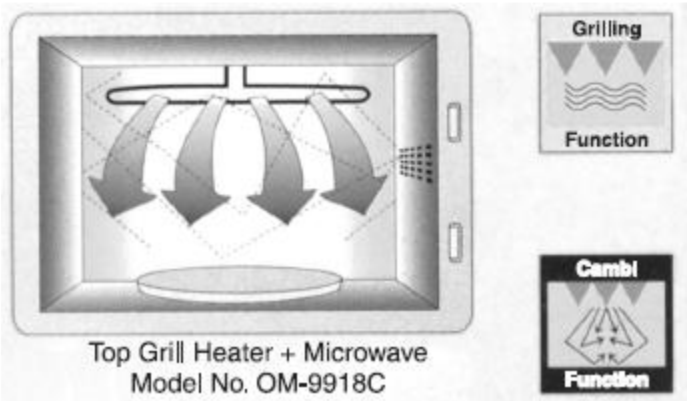


Fig. 50.9 In combo mode of cooking, the microwave along with the grill can be used simultaneously to cook faster for perfect browning and crispy texture.

In multi-grill models, the 3D power system, Fig. 50.10 comprises of heaters placed at different places (top, bottom and rear) in the oven, together with a convection fan forming a cube heater. This not only cooks food evenly but also saves time. It allows you to do your own heat setting and helps to get the perfect crispy base, overcrispy top, or both at the same time.

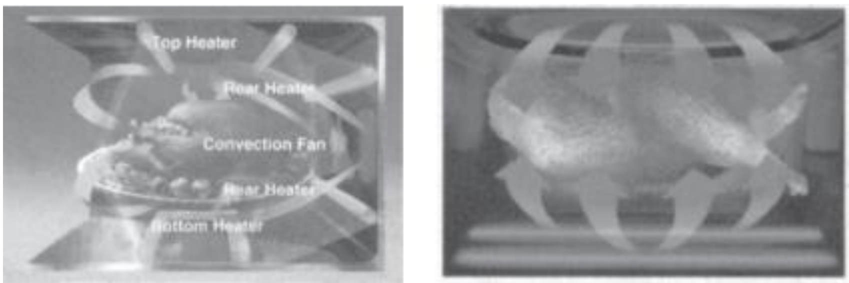


Fig. 50.10 3D power convection cube heater (b) and multi grill system (R)

MICROWAVE COOKING

Rapid microwave cooking provides excellent quality of food. This quick method of cooking using a minimum quantity of water helps *retain* most of the nutrients. In addition microwaving preserves natural flavours, while *enhancing* the colour and texture of the food.

Microwaves are non-ionising, high-frequency, short wavelength electromagnetic waves. Microwaves are *attracted* by the moisture in the food placed within the oven for cooking. The microwaves then *penetrate* the food surface, causing the moisture molecules to vibrate. This vibration *generates* heat which then cooks the food by conduction. The food is cooked from the outer surface to the inner core.

FEATURES DIAGRAM

Heating through microwaves has the following *advantages* over conventional electrical heating systems

1. Quicker heating saves time.
2. System is very clean.
3. Preserves natural flavours.
4. Retains most of the nutrients.
5. Enhances the colour and texture of food.
6. System can be combined with a conventional heating process.

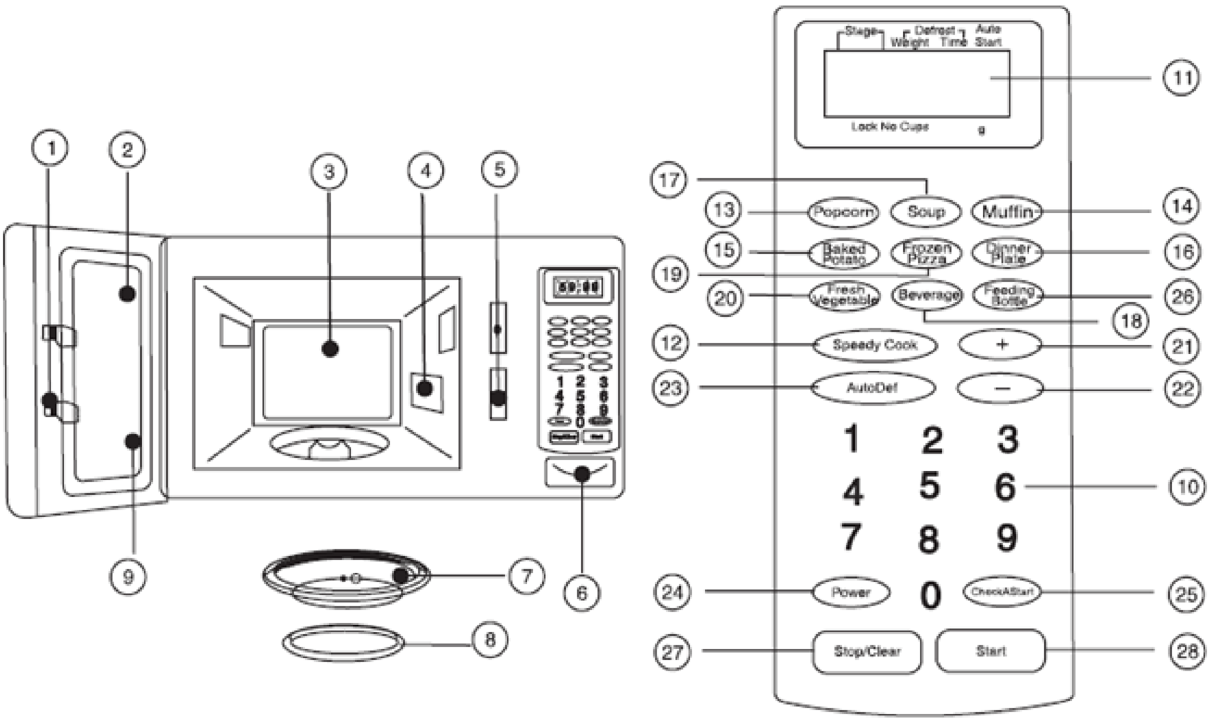


Fig. 50.11

1. Door Latch	–	When the door is closed it will automatically *lock shut*. If the door is opened while the oven is operating the magnetron will be *automatically switched off*.
2. Door Seal	–	The door seal maintains the microwave within the oven cavity and *prevents microwave leakage*.
3. Oven Cavity.		
4. Spatter Shield	–	Protects the microwave outlet from splashes of cooking foods.
5. Safety Interlock System	–	*Prevents the oven from operating while the door is opened.*
6.Door Release Button	–	Pushing this button stops oven operation and *opens* the door.
7. Glass Cooking Tray	–	Made of special *heat resistant glass*. The tray must always be in proper position before operating. *Do not cook food directly on thetray.*
8. Roller Guide	–	*Supports* the glass cooking tray.
9. Door Screen	–	*Allows viewing of food*. The screen is designed so that light can pass through, but not the microwaves.
10. Time Set Pad	–	Used to set the *cooking time* and the *present time*.
11. Display	–	Cooking time, power level, present time *displayed*.
12. Speedy Cook	–	Touch to set any desired *reheat settle*
13. Popcorn	–	Used to *cook* popcorn.
14. Muffin	–	Used to *cook* muffin.
15. Baked Potato	–	Used to *bake* potatoes.
16. Dinner Plate	–	Used to *reheat* dinner plate.
17. Soup	–	Used to *reheat* soup.
18. Beverage	–	Used to *reheat* beverage.

19. Frozen Pizza	–	Used to *reheat* frozen pizza.
20. Fresh Vegetable	–	Used to *blanch* fresh vegetable.
21. More	–	Used to *add* on one touch cooking.
22. Less	–	Used to *remove* one touch cooking.
23. Auto Defrost	–	Used to *defrost* foods.
24. Power	–	Used to set *power level.*
25. Clock/A. Start	–	Used to set clock & used to set *auto settle.*
26. Feeding Bottle	–	Used to *sterilize* bottle.
27. Stop/Clear	–	Used to *stop* the oven operation or to *defrost.*
28.Start	–	Used to start a *selected operation.*

WIRING INSTRUCTIONS

The wires in this mains cord are coloured in accordance with the following code.

Green : Earth

Black : Neutral

Red : Live

As the colours of the wires of the mainscord of this appliance may not correspond with the coloured marking identifying the terminals in your plug, proceed as follows : The wire which is coloured *green* must be connected to the terminal in the plug which is marked with the 'E' or by the earth symbol or green. The wire which is coloured *black* must be connected to the terminal which is marked with the letter 'N' or coloured black.

The wire which is coloured *red* must be connected to terminal which is marked with the letter 'L' or coloured red. Ensure *proper wiring* of the socket which should be of *15 Amps* capacity. Line terminals should confirm to the above.

Warning : This appliance must be *earthed properly*.

SAFETY INSTRUCTIONS

Listed below are, as with other appliances, certain rules to follow and safeguards to assure best performance from this oven :

1. Do not use the oven for drying clothes, paper or any other *nonfood item.*
2. Do not use the oven *without* food items, this could damage the oven and may cause smoke emission.
3. Do not use the oven for *storage* of papers, cookbook, cookware, etc.
4. Do not operate the oven without glass tray. *Be sure it is properly placed on the rotating base.*
5. Ensure *removal* of caps or lids prior to cooking when you cook food sealed in bottles.
6. Do not put *foreign material* between the oven surface and door which could result in excessive leakage of harmful microwave energy.
7. Do not use *recycled paper products* for cooking. They may contain impurities which could cause sparks and/or fires when used during cooking.
8. Use *recommended & commercially packaged* popcorn. Microwave popped corn produces a lower yield than conventional popping, there will be a number of unpopped kernels. Do not use oil unless specified by the manufacturer.
9. Do not pop popcorn longer than the manufacturer's directions *(popping time* is generally below 3 minutes). Longer cooking does not yield more popped corn, it can cause scorchings and fire. Also, the cooking tray can become too hot to handle or may break.

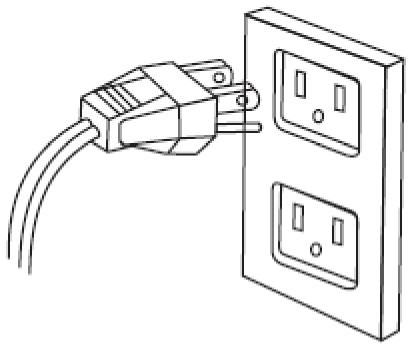


Fig. 50.12 Properly polarised and grounded outlet

10. Do not cook any food surrounded by a membranes such as egg yolks, potatoes, chicken livers, etc., *without piercing them.*
11. Should the microwave oven emit smoke indicating a fire, *keep the oven door shut*, switch the appliance off and disconnect the mains cord from the outlet.
12. When flammable food containers are used in the oven (e.g. packet popcorn) be sure to *check the cooking process frequently* to check for fire.
13. Always *stir and/or shake* the containers of baby foods prior to testing their temperature and serving the contents.
14. Always *test the temperature* of food or drink which has been heated in a microwave oven before serving, especially to children or elderly people. This is important

because things which have been heated in a microwave oven keep on getting hotter even though the microwave oven cooking has stopped.

MICROWAVE COOKWARE

A wide variety of cookware can be used in the microwave oven, all of which are termed *microproof* to indicate that it is safe and recommended for microwave cooking. Most materials are microproof for at least a limited amount of cooking time. The principal exception to this rule is metal. *Unless specifically approved, containers made of metal, even partially, are never to be used in the microwave oven.* Not only does metal reflect microwaves, preventing them from entering a food, but it can also cause arcing, a condition where a static charge is conduced between the metal container and walls of your oven. Arcing is not dangerous to you, but it can deface your oven.

When selecting *containers* to cook in, be sure to consider what size will fit in your oven. After all, the cavity space of this oven is much smaller than the 2 to 2.5 cubic feet you are probably used to with your conventional oven. To familiarise yourself with the capacity of your microwave oven, it may be best to experiment with different container shapes and sizes.

OPERATING PROBLEMS AND SOLUTIONS

You can often correct *operating problems* yourself. If your microwave oven fails to work properly, locate the operating *problems* in the chart below and try the *solutions* marked for each problem. If the microwave oven still does not work properly contact the nearest Service Centre, which are fully equipped to handle all of your service requirements.

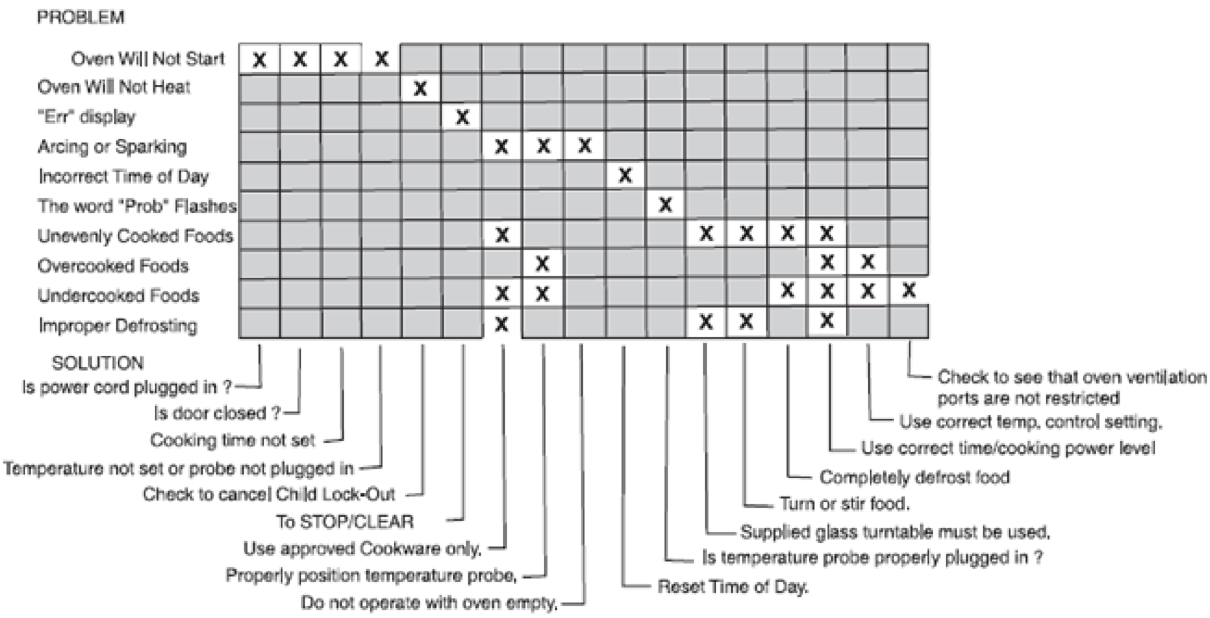


Fig. 50.13

CARE AND CLEANING

Wipe the oven inside and outside with a *soft cloth and mild detergent solution*. Then rinse and wipe dry. This should be done on *a weekly basis*—more often if needed. Never use rough powders or pads.

The inside oven top can be *gently wiped* in place. Excessive oil spatters on the inside top will be difficult to remove if left for many days.

Wipe spatters with a *wet paper towel* especially after cooking chicken or bacon.

IMPORTANT : *If interior of the oven is not kept clean, the stirrer fan assembly will accumulate grease and food stains which will shorten the life of the stirrer fan assembly parts.*

REMOVABLE PARTS

The following parts may be removed as described. They should be washed in warm (not hot) water with a mild detergent and a soft cloth. *Once they are clean, rinse them well and dry with a soft cloth.*

Never use rough cleaning powders, steel wool or rough cleaning pads.

1. After each use of *temperature probe*, it must be removed from the socket. Use pot holder, as the wire, plug and sensor section may be hot. Wipe food or liquid from the sensor with a soft damp cloth.
2. The *glass turntable* may be removed for cleaning at the sink. Wipe up spillovers with a paper towel or cloth before removal of the glass turntable. Be careful not to chip or scratch the edges of the glass turntable, as this may cause the glass turntable to break during use.
3. The turntable roller rest and oven cavity bottom should be cleaned regularly to avoid excessive noise. Simply wipe the *oven bottom surface* with mild detergent water and dry. The *turntable roller* rest may be washed in warm (not hot) water with a mild detergent and a soft cloth. Cooking vapours collect during repeated use, but in no way affects the oven bottom surface or roller rest wheels.

SPECIAL CARE

For best performance and safety, the inner door panel and the oven front frame should be free of food or grease build-up. Wipe both often with a mild detergent. Then rinse and wipe dry. *Never use rough powders or pads.*

After cleaning the control panel, *touch STOP/CLEAR*. This will clear any entries that might have been entered accidentally while cleaning the panel.

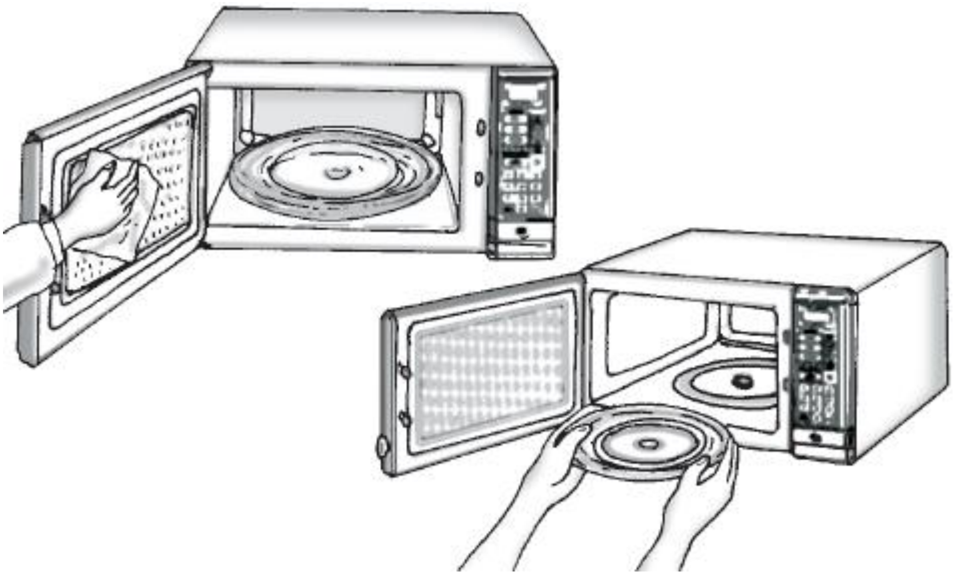


Fig. 50.14

METAL RACK CARE

1. The metal rack may get hot during cooking. Pot holders may be needed to remove Rack after cooking.
2. Remove metal rack from oven when not being used for whole meal cooking.
3. Do not use browning dishes on metal rack.
4. Do not run the oven empty with the metal rack in it.
5. Do not use foil or metal containers on the metal rack.

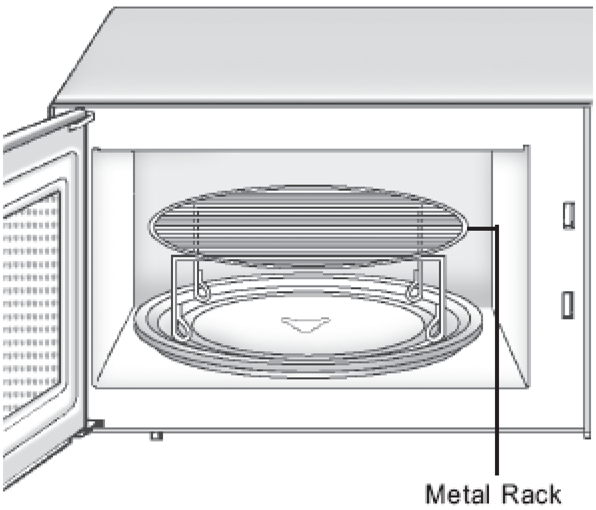


Fig. 50.14

EXERCISES

Descriptive Questions

1. Explain the significance of microwaves.
2. Briefly explain transit time effect.
3. Explain the working of a multicavity magnetron.
4. What is the practical application of waveguides?
5. Explain the following :
   a. Wave reflector system
   b. Dual-wave emission system
   c. Fail safe system in microwave ovens
   d. 3D microwaves
6. Draw the block diagram of a microwave oven. Briefly explain each block.
7. What are the types of microwave ovens?
8. What are the advantages of microwave cooking?
9. With the help of a suitable block diagram explain the working of an LCD timer with alarm.
10. Briefly explain the working of a single-chip microcontroller.

Fill in the Blanks

1. The microwave oven uses microwaves of frequency ________.
2. Electricity applied to the magnetron is used to create ________.
3. 3. The wave reflection system and dual-wave emission system ensure that food is ________.
4. All microwave ________ remains inside the oven.
5. Waveguides are the ________ hardware.
6. Microwave energy is carried through the waveguide by ____________________ along its inside walls.
7. The microwave oven is a ________ device.
8. The anode of a magnetron is at a high ________ potential.
9. The thermal protector in a microwave oven is a ________ thermistor.
10. In LCD timer with alarm the keypad is the input and both the digital ________ and ________ are the output devices.
11. While the microwave oven remains popular there is now a greater demand for ________.
12. Microwave cooking ________ most of the nutrients and ________ the colour and texture of food.

ANSWERS

Fill in the Blanks

1.	2.4 GHz
2.	microwave energy
3.	uniformly
4.	activity
5.	inter connecting
6.	reflection

7. fail safe
8. negative
9. PTC
10. display; alarm buzzer
11. combination ovens
12. retains; enhances

CHAPTER 51

WASHING MACHINES

From the first washing tool, a broom with four fingers at the bottom to move the clothes around the bucket, to the modern fully automated ones, washing machines have come a long way. Technological advancements have brought about metamorphic changes in washing machines. From manual washing machines requiring hot water soak before dirty clothes got churned in machines to semi-automatic with spin dry facilities to fully automatics and now recently introduced Fuzzy logic the concept of machine wash has totally changed. Multinational companies (MNCs) dealing in electronics have launched new and modern techniques in washing machines.

In today's high-stress life a washing machine has become a household necessity. Washing machines are gradually emerging as an omnipresent dhobi in Indian homes. Fortunately the choices are many. Not only in colour, design and features but also in prices. Of course the market is flooded with washing machines offering a range of operations. From semi-automatic to fully automatic, from top-load to tumble wash, these state-of-the-art machines promise multifarious, user friendly features at down to earth prices.

ELECTRONIC CONTROLLER FOR WASHING MACHINES

The task here is simply to identify the *input and output devices* used in electronic washing machines and to construct a block diagram showing their connections to the controller. Detailed information about the characteristics of sensors and actuators can be added at a later stage.

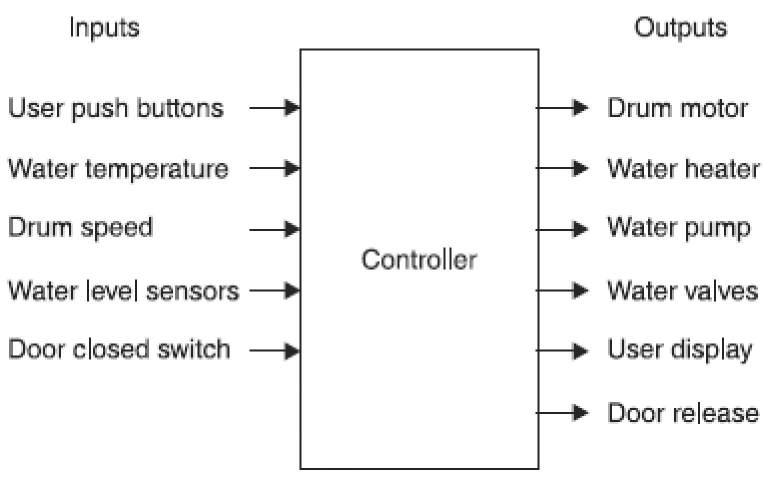


Fig. 51.1 Inputs and outputs in an electronic washing machine

The block diagram in Fig. 51.1 shows a possible representation of the system. There are many acceptable ways of representing the system. It would, for example, be possible to consider the *display* to be internal to the controller and therefore not show it separately. Similarly *clock circuitry* used to time the operation of the machine is considered here to be contained within the controller. It could equally well be considered as an external component.

The block diagram is a good starting point for the generation of the specification since it shows very clearly the structure of the complete system. The block diagram makes no assumptions of the *form* of the controller. It could be implemented using an electromechanical timer, or a microcomputer, or a range of other technologies.

Many modern washing machines now use microcomputer to control their various functions, replacing the electromechanical controllers used in earlier models. Clearly it is not practical to consider all aspects of such a system, but it is instructive to look at some elements of the design.

At various stages of the *washing cycle* the drum is required to rotate at different speeds. These include: a *low speed* of about 30 revolutions per minute (rpm) while clothes are washed: an *intermediate speed* of about 90 rpm while the water is pumped out and a *high speed* of either 500 or 1000 rpm to spin dry the clothes. Let's consider how the microcomputer should control the speed of the motor.

Since a domestic washing machine is a very high-volume product, the design should attempt to minimise the amout of hardware required. This necessitates a close look at the choice of sensors and actuators to select low-cost items. Our first decision must be whether the system will be open loop or closed loop. Since although an *open-loop* system is theoretically possible using a synchronous motor the cost of such a system for high-power variable-speed applications is prohibitive. The system will therefore be *closed loop* using a motor to drive the drum and some form of sensor to measure its speed.

One of the simplest methods of speed measurement is to use a counting technique illustrated in Fig. 51.3. It uses a fixed inductive sensor to produce a pulse each time it is passed by a magnet which rotates with the drum. This produces *one pulse per revolution* of the drum which can be used to determine its speed.

The speed of the motor will be controlled by the power dissipated in it. The simplest way of *speed control* is to use a triac. The power could be controlled by some form of electronic circuitry, but the hardware requirement can be reduced if the microcomputer controls the power *directly* by firing the triac at an appropriate time during its cycle. To do this the controller must detect the *zero crossing* of the ac supply. This will require circuitry to detect the crossing point while protecting the processor from high voltages. A block diagram of the system is shown in Fig. 51.2.

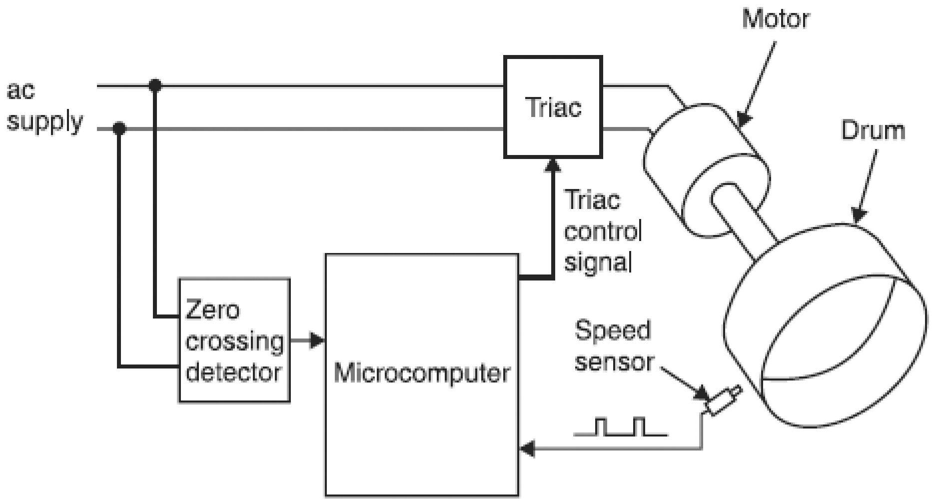


Fig. 51.2 Washing machine control

At any time in the washing cycle the program determines at what speed the drum should rotate. From a knowledge of the required speed and the actual speed as obtained above, the controller can determine whether to increase or decrease the power dissipated in the motor.

The motor power is determined by the timing of the triac firing pulse. If the triac is fired at the beginning of each half of mains cycle it will remain on for the remainder of the half cycle and the motor will operate at full power. *The longer the processor waits before firing the triac, the less will be the motor power.* The processor thus varies the delay time with respect to the zero crossing point of the mains by an appropriate amount to increase or decrease the power in the motor as determined by the difference between the actual and required speeds.

This method of controlling the motor speed is very processor intensive. It consumes a large amount of processor time and will require a considerable amount of effort in writing and developing the software. However, this approach uses very little hardware and is thus very attractive for such a high-volume application.

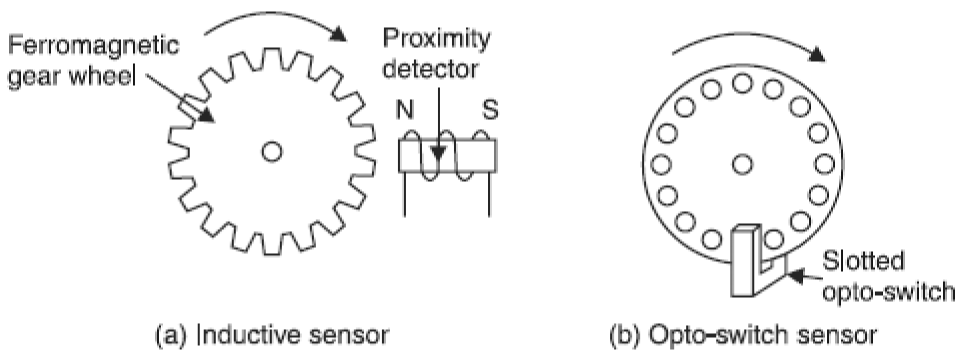


Fig. 51.3 Displacement sensors using counting

WASHING MACHINE HARDWARE

A *system* is an assembly of components united by some form of regulated interaction to form an organised whole. We will examine a microcomputer system, using a washing machine control as an example. The input peripherals consist of (Fig. 51.4).

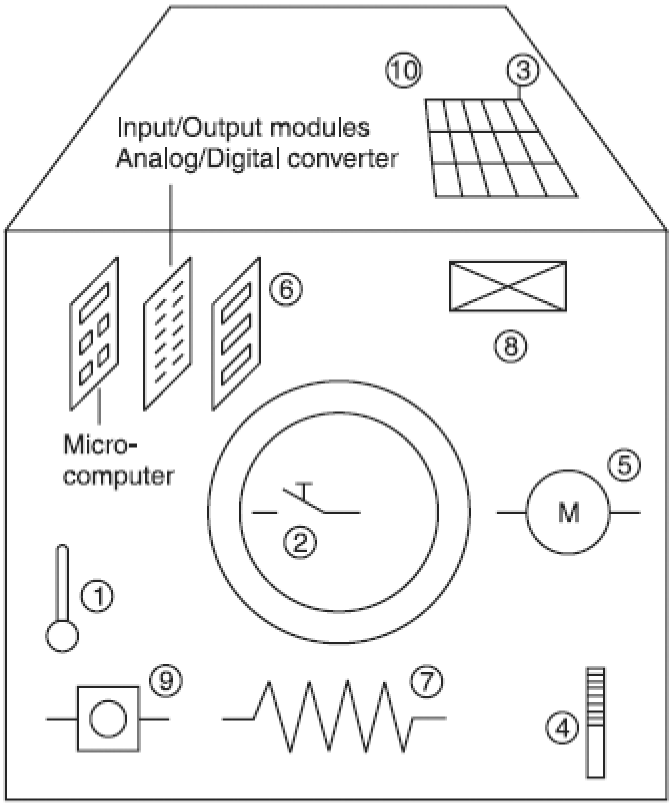


Fig. 51.4 Washing machine—hardware

1. temperature sensor which senses the washing water temperature. (The analog/digital converter changes the analog values to binary numbers).
2. safety cut-out switch.
3. keyboard for program selection.
4. water level gauge.
5. motor for washing drum.
6. power switches for motor, heater, etc.
7. heater for washing water.
8. water inlet valve.
9. water suction pump.
10. control lamps and indicators.

The units listed above i.e. the washing machine as well as its mechanical components, electrical units and electronic components are known as *hardware.*

WASHING CYCLE

The push-button keyboard enables the desired program to be selected. The control—the microcomputer— checks firstly that the safety cut-out is in the ON position. The water is then admitted (valve opened) and the water level is constantly monitored. When the required quantity of water has been provided the valve closes. The water temperature is measured and the heater is switched until the water reaches the required temperature. In the meantime, the washing powder is admitted from a container and the hardness of water is noted, at the same time the drum motor is switched on so that the dirty washing is *evenly* moved through the water. After the required time has elapsed, according to the selected program, the motor is switched to *high speed spinning* and the suction pump is switched on to remove the washing water and the rinsing water to waste. At the end of the *washing cycle* the machine switches off and provides a signal to indicate this.

HARDWARE AND SOFTWARE DEVELOPMENT

We will now examine how a system is developed. The example used for this is, of course, a simple washing machine control. The development will follow the broad pattern shown in Fig. 51.5.

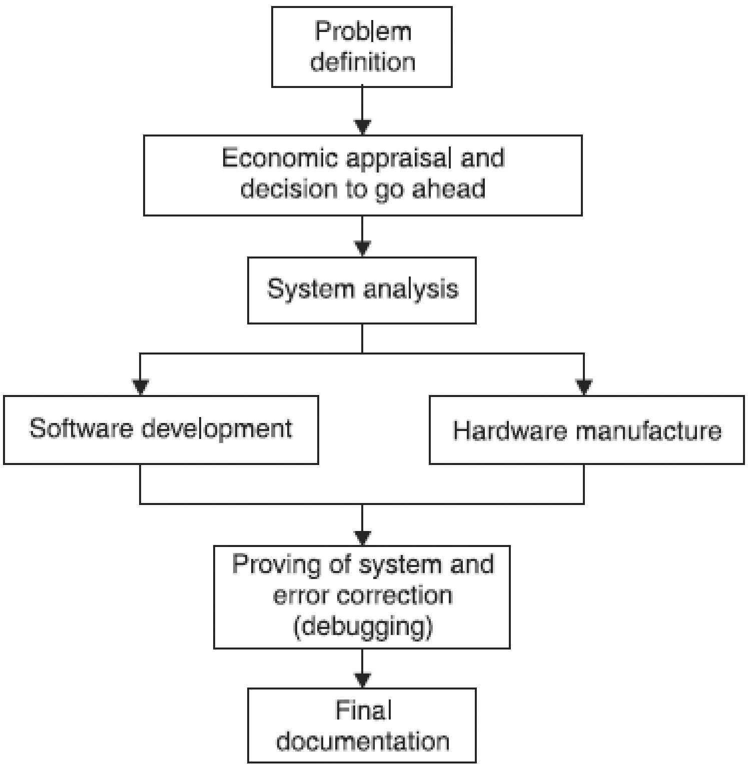


Fig. 51.5 Developing the system for washing machine control

The problem definition is based on the requirements of the specification. It is also necessary for the redesign of the existing unit. It is a means of determining what a system's performance is capable of and what is required from it.

Data flow charts are used to identify all the hardware elements of a system at this stage for a general broad picture of the structure of the installation.

Program flow charts permit the costs of the necessary software to be established in the development stage and represent useful aids for the designer.

The decision to go ahead with the developments of a system is governed by *economic appraisal* and *technical feasibility* of the plan. To establish these criteria the required operating speed, memory storage capacity and costs of the component parts of the system must be determined. Subsequently the structure of the problem is analysed and the final production costs deduced.

There are two alternative approaches for *hardware development.* On one hand, a *universal system* may be considered which has not been designed to cope with any one specific problem. On the other hand a *specially designed system* may be decided upon in which the components used are specially selected for their suitability to deal with the problem under consideration. Such optimization is generally not possible when standard systems are employed.

For *software development* a detailed program sequence plan must first be established. This is then written in the appropriate code and fed into a computer or into a development system. The program is then *translated* into the language required by the machine and a *simulation* of the operation sequence is carried out. Any errors found in the program are corrected (this is known as *debugging)* and the software is then available for use.

After the hardware and software has been developed the system is *tested.* An examination is carried out to determine whether the system can satisfy all the demands which may be put upon it, i.e. "Can the machine perform every function which may be required from it?" It is not now a question of testing the program (this has already been done during the program development) but *the system is now under scrutiny.*

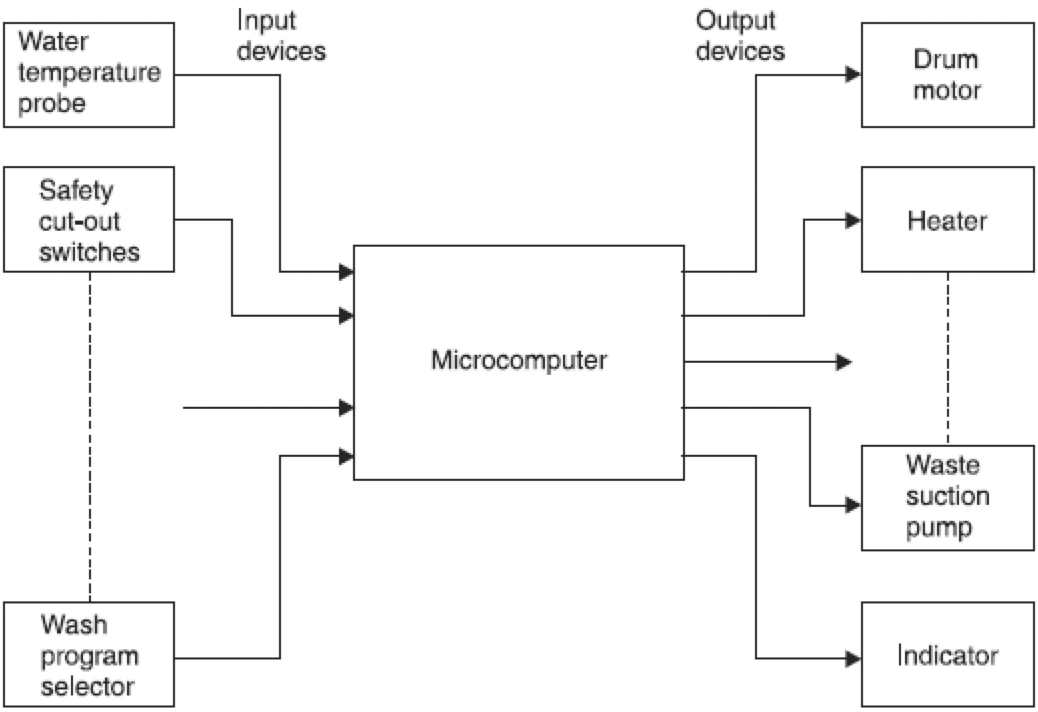


Fig. 51.6 Data flow chart for a washing machine control

TYPES OF WASHING MACHINES

Washing machines are mainly of three types, namely washer, semi-automatic and automatic. *Washers* are single tub machines that only wash. Since washers don't have the facilities for drying the clothes, these cost less than semi-automatic and fully automatic machines.

In *semi-automatic* machine, Fig. 51.8, the controls are not fully automatic and manual intervention is required.

In *fully automatic* machines, Fig. 51.9, no manual intervention is required during the washing process. For automatic machines, programs have to be selected and set by the user prior to the start of washing cycle. Sensors sense the wash load and decide the program ideal for washing the clothes, water level, time required to wash, number of rinses and spins, type of fabric etc.

Although *washer dryer* (semi-automatic) machines don't operate with the efficiency of *stand alone* washing machines, they offer enormous space saving. However, you have to drain all the soap water before drying. Also, you can't wash and dry at the same time and the drying performance is inferior to that of stand alone machines. But then washer-dryers cost less and allow you to wash and dry your clothes without having to reset the machines.

FUZZY LOGIC WASHING MACHINES

Fuzzy logic washing machines are gaining popularity. These machines offer the advantages of performance productivity, simplicity, and less cost. Sensors continually monitor varying conditions inside the machine and accordingly adjust operations for the best wash results. As there is no standard for fuzzy logic, different machines perform in different manners.

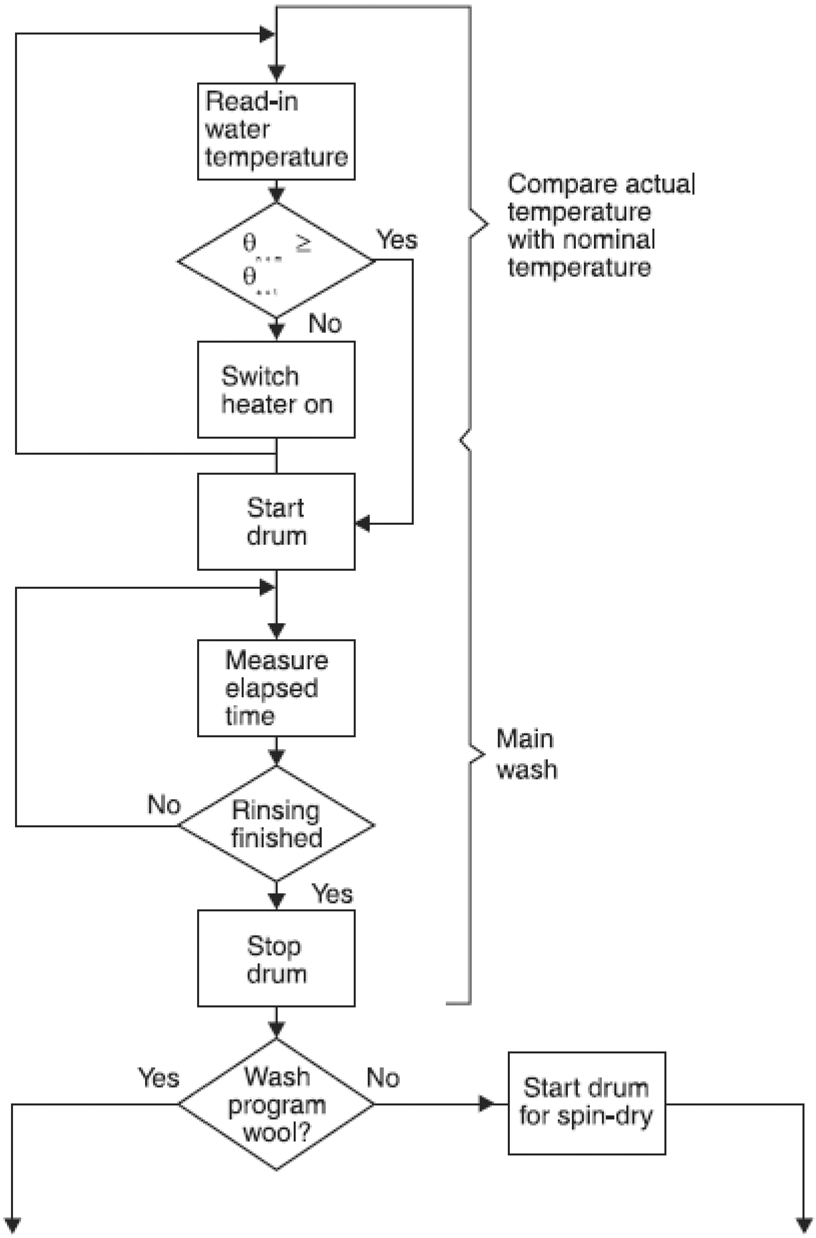


Fig. 51.7 Program flow chart for a washing machine control

Typically, fuzzy logic controls the washing process, water intake, water temperature, wash time, rinse performance and spin speed. This optimizes the life span of the washing machine. More sophisticated washing machines weigh the load (so you can't overload the washing machine), advise on the required amount of detergent, assess cloth material type and water hardness, and check whether the detergent is in powder or liquid form. *Some machines even learn from past experience, memorising programs and adjusting them to minimise running costs.*

The diagnostic fault-finding system displays a *fault code* if any problem arises. You can then convey this fault code to the service centre thus ensuring that the repair technician reaches with the right parts to fix it without delay.

Machines with *fuzzy logic microprocessors* can be updated as and when a new technology or program comes up. Several models of internet-enabled washing machines have been launched. When the network home becomes a reality these machines will allow downloading of new programs and remote fault diagnosis over the direct internet connection.

Most fuzzy logic machines feature *one-touch control*. Equipped with energy saving features, these machines consume less power, and are worth paying extra for if you wash full loads more than three times a week. In-built sensors monitor the washing process and make corrections to produce the best washing results. In some machines a *tangle sensor senses* where the clothes are tangled and takes corrective action by adjusting the water current, so the clothes don't tangle further and are cleaned better.

Fig. 51.8 A typical semi-automatic washing machine (top loading)

High-end machines have a *suds-free system* including a pressure sensor to detect extra suds in washing if you have used a large amount of detergent. The washing machine drains water together with the detergent and then refills with minimum water to restart. These machines cost more than regular models. The *foam suppression feature* detects whether too much foam is present during wash and accordingly it either reduces the agitation or adds an extra rinse.

Fuzzy logic checks for the extent of dirt and grease, the amount of soap and water to add, direction of spin and so on. The machine *rebalances* washing load to ensure correct spinning. Else it reduces spinning speed if an *imbalance* is detected. *Even distribution* of washing load reduces spinning noise.

Neuro-fuzzy logic incorporates optical sensors to sense the dirt in water and fabric sensor to detect the type of fabric and accordingly adjust wash cycle.

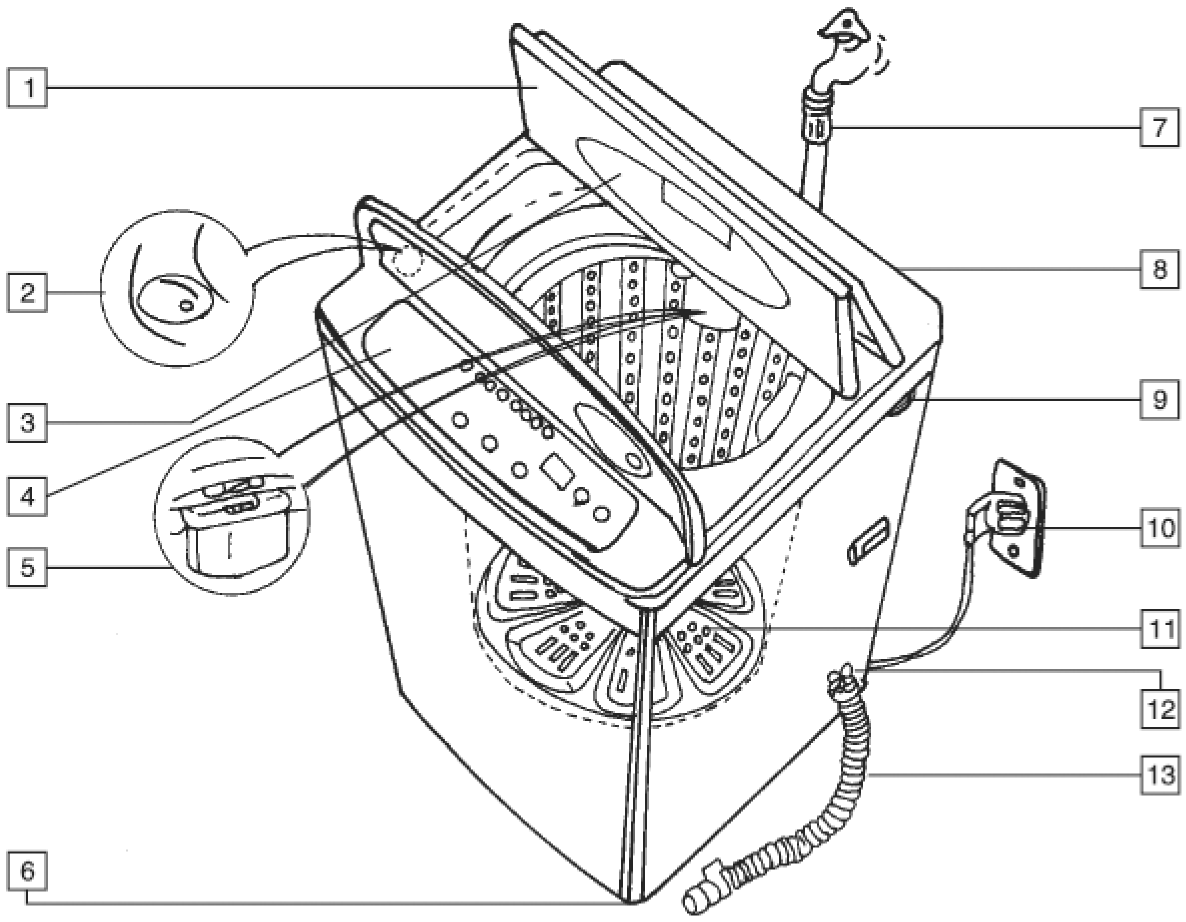
MISCELLANEOUS FEATURES

The controls and features of a typical *top loading* washing machine are shown in Fig. 51.10. The controls of a typical *front loading* washing machine are given in Fig. 51.11.

Fig. 51.9 A typical fully-automatic washing machine (top loading)

Washing machines incorporate a tub with heating element and something to rotate or scrub the clothes in the drum. Once the water and detergent are added mechanical action begins to *soak and agitate* the clothes. Fuzzy logic electronics intelligently improves the wash performance in washing machines.

1. **Capacity :** The capacity of a washing machine is expressed in terms of the *wash load*, which in turn depends on the type of fabric. It is expressed in kg. The maximum load for the washer is the amount that will move freely in the wash tub. Indicative range of weights of some commonly washed clothes is given in Table 51.1.
   A higher capacity machine offers the covenience of washing more clothes at one go but consumes more power. Smaller capacity machines wash fewer clothes and consumes less power, but these machines can easily fit in a limited space.
2. **Wash programs :** High-end washing machines feature different wash programs to suit different types of clothes. The program includes *regular* for normal wash, *gentle* for delicate clothes and *tough/hard* for rugged clothes. In addition, you are able to select the temperature of wash and the number of runs for better cleaning. The number of cycles specifies the number of *preset programs* available on the machine. This is important for clothes that require different temperatures.
3. **Spin Speed :** The higher the spin speed, the dryer the clothes at the end of the *washing cycle* and hence the shorter the drying time in the tumbler dryer. Thus a high spin speed results in less *washing time.* Some machines spin at more than 1000 rpm, some machines spin as fast as 7000 rpm during *drying cycle.*



1. **Lid**
2. **Bleach inlet**
3. **Transparent window**
   You can see the laundry being washed through this window.
4. **Front control panel**
   Open the panel lid, and then choose the function you want to use.
5. **Softener case**
6. **Adjustable legs**
   Adjust the length of the legs when installing the washer.
7. **Water supply hose**
8. **Lint filter**
   Lint will collect in the lint filter pouch during washing.
9. **Drain hose hook**
10. **Power cord**
11. **Pulsator**
12. **Drain hose clamp**
13. **Drain hose**

Fig. 51.10 Controls and features of a typical top loading washing machine

4. **Washing Technique :** In some machines a *pulsator disk* (Fig. 51.12) at the bottom, circulates water upwards in large circles while rotating, providing better and gentler cleaning of clothes. In the *agitator wash technique* a rod with fins (Fig. 51.13) is used at the centre of the washing machine. A rubbing action squeezes the dirt out of clothes. But it restricts the space and the clothes tend to get entangled. The *tumble wash technique* is used in front loaders. A steel drum rotates along a horizontal axis and the clothes rub against its metal surface due to centrifugal action. The cleaning is, of course, superior but there is a risk of ruining gentle fabrics.

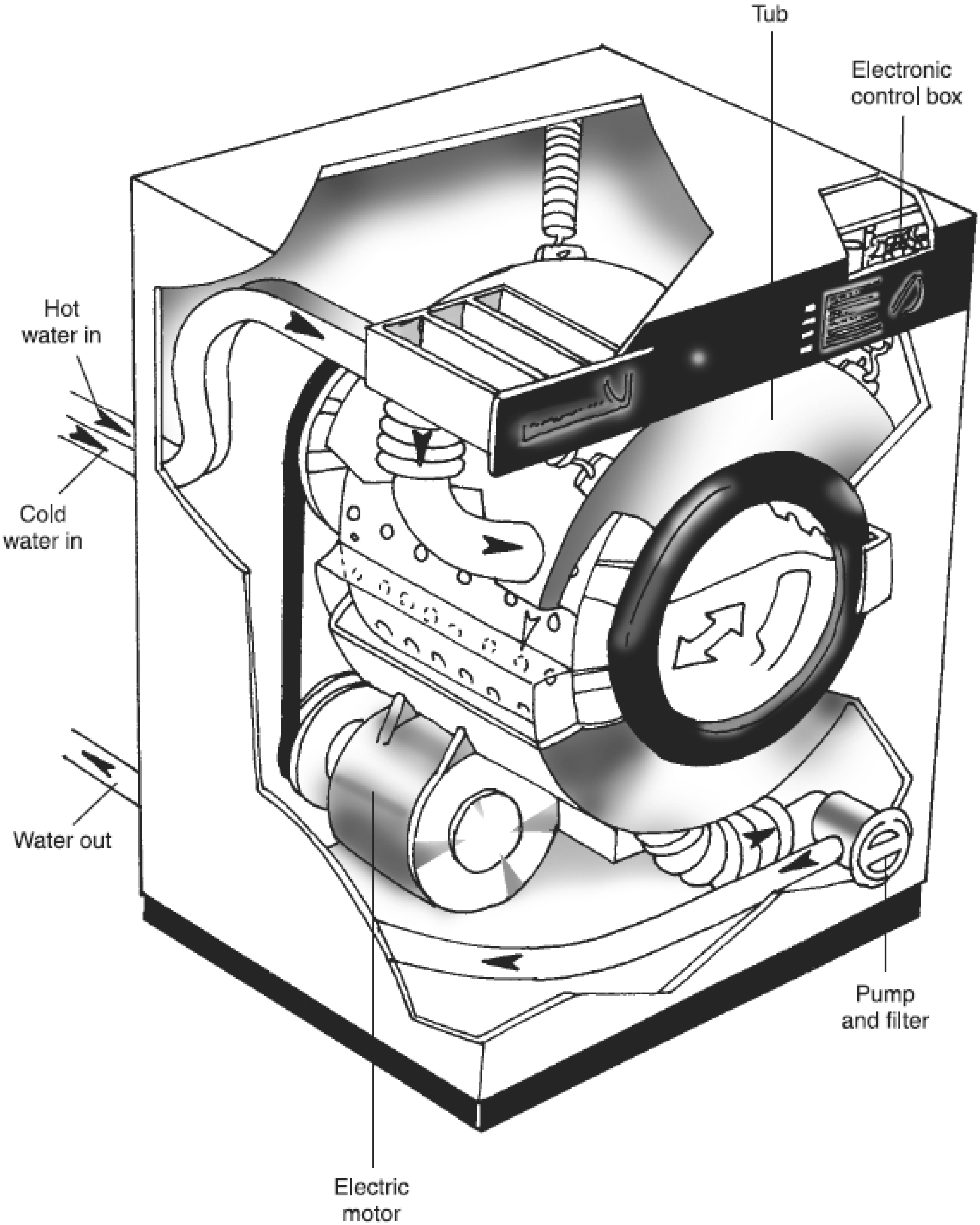


Fig. 51.11 Controls and features of a typical front loading washing machine

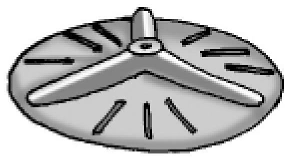
In *LG punch + 3 technique* the water punch propels the detergent rich water vertically into every thread of the fabric. The action is supported by three *mini pulsators* which work with the main pulsator to generate powerful micro water-eddies. The mini pulsators rotate in the *opposite* direction to the main pulsator. This helps in reducing entanglement of clothes, resulting in less wear and tear and better wash technology.

5. **Loading the machine :** *Top loaders* (Fig. 51.10) allow you to easily remove clothes, without having to bend even during power failure. These are compact and require

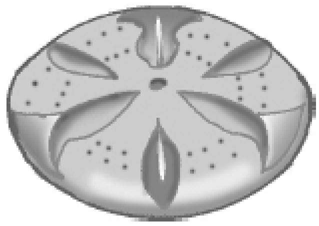
normal detergents. You can add clothes even during the wash cycle. The larger the porthole, the more convenient the loading and unloading. Most top loading machines have an agitator.

Front loaders (Fig. 51.11) are usually more expensive than top loaders as these incorporate heftier motors and suspensions. However, these machines consume less water and dry clothes much faster, thereby reducing energy bill. The hot wash option allows better cleaning. You cannot open a front loader midway through a wash cycle. You need to use detergents producing less lather and if the power fails you can't open the door due to water in the drum. Also you need to leave room for door opening/closing on the front side.

6. **Automation :** On fully-automated washing machines you don't need to wet your hands, just put in the wash load, turn the machine on and wait for it to finish washing and drying. *Automatic machines require a dedicated running water supply from a tap.* A single tub carries out all the actions. The washing machine does washing, rinsing and drying and beeps when it is through with all the tasks.



Conventional pulsator



Larger boomerang pulsator

Fig. 51.12 The pulsator disk

Fig. 51.13 The agitator rod

Fig. 51.14 LG punch +3 technique

Table 51.1 Indicative Range of Weights of Some Commonly Washed Clothes

Clothes	Approximate weights
Shirt	200 gm – 300 gm
Trousers	350 gm – 500 gm
Pyjama	300 gm – 400 gm
Kurta	200 gm – 300 gm
Lungi	200 gm – 300 gm
Vests	75 gm – 100 gm
Underwear	100 gm – 150 gm
Socks	50 gm – 75 gm
Dhoti	200 gm – 250 gm
Saree	300 gm – 500 gm

Blouse	100 gm – 150 gm
Petticoat	350 gm – 450 gm
Undergarment	75 gm – 150 gm
Nightgown	200 gm – 300 gm
Salwar Suit	350 gm – 500 gm
Frocks-Baby	200 gm – 300 gm
Frocks-Girl	350 gm – 450 gm
Skirt	200 gm – 300 gm
Sweater	500 gm – 800 gm
Bedsheet Double	1000 gm – 1200 gm
Bedsheet Single	400 gm – 600 gm
Towel large	700 gm – 1000 gm

On semi-automatic machines you have to *manually* transfer the clothes from the washer to the dryer. Semi-automatic machines featuring microprocessor based controls with feather-touch buttons consume less power and are *preferable where running water is not available.*

EXERCISES

Descriptive Questions

1. List the inputs and outputs of a washing machine.
2. Explain washing machine control.
3. Explain the sequence of operations in a wash cycle.
4. What are the different types of washing machines?
5. What are the features available with washing machines?
6. Differentiate between :
    a. A top loader and a front loader
    b. A semi-automatic and a fully-automatic washing machine.
    c. A pulsator and an agitator.
7. Describe the working of a neuro-fuzzy washing machine.
8. What additional features are available with washing machines with fuzzy logic as compared to those available with fully-automatic washing machines?

Fill in the Blanks

1. At various stages of the wash cycle the drum is required to rotate at ——————————
2. One of the simplest methods of speed measurement is to use a ———————— technique.
3. The washing machine as well as its mechanical components, electrical units and electrical components are called ————————
4. Washers are ———————— machines.
5. In semi-automatic washing machines ———————— intervention is required.
6. In fully-automatic machines ———————— have to be selected and set by the user prior to the start of the———————— cycle.
7. The diagnostic fault finding system displays a ———————— if any problem arises.
8. Most fuzzy logic machines feature ———————— control.
9. Fuzzy logic machines ———————— washing load to ensure correct spinning.
10. In the agitator wash technique a ———————— with fins is used at the centre of the washing machine.
11. The tumble wash technique is used in ————————
12. Sensors continually monitor ———————— inside the machine and accordingly adjust operations for the best wash results.

ANSWERS

Fill in the Blanks

1.	different speeds
2.	counting
3.	hardware
4.	single tub
5.	manual
6.	programs; washing
7.	fault code
8.	one touch
9.	rebalance
10.	rod
11.	front loaders
12.	varying conditions

CHAPTER 52

AIR CONDITIONERS AND REFRIGERATORS

A typical person in modern society may spend upto 90% of each day indoors. It is not surprising, therefore, that providing a healthy, comfortable indoor environment has become a major factor in our economy.

To an average person, air conditioning simply means "the cooling of air". This definition is neither sufficiently useful nor accurate. To be more precise, air conditioning is the process of treating air in an internal environment to establish and maintain required standards of temperature, humidity, cleanliness, and motion. Most air conditioning systems are used for either human comfort or for process control.

AIR CONDITIONING

Air conditioning is the process of treating air in an *internal environment* to establish and maintain required standards of temperature, humidity, cleanliness, and motion. This is how each of these conditions is controlled:

1. *Temperature :* Air temperature is controlled by heating or cooling the air. Cooling technically means the *removal* of heat, in contrast to heating, the *addition* of heat.
2. *Humidity :* Air humidity, the water vapour content of the air, is controlled by adding *(humidification)* or removing *(dehumidification)* water vapour from the air.
3. *Cleanliness :* Air cleanliness or air *quality* is controlled by either *filtration*, the removal of undesirable contaminants using filters or other devices or by *ventilation*, the introduction of outside air into the space which dilutes the concentration of contaminants. Often *both* filtration and ventilation are used in an installation.
4. *Motion :* Air motion refers to air velocity and to where the air is *distributed.* It is controlled by appropriate air distributing equipment.

Sound control can be considered an *auxiliary function* of an air conditioning system even though the system itself may be the cause of the problem. The air conditioning equipment may produce excessive noise requiring additional sound attenuating (reducing) devices as part of the equipment.

The above description does not imply that every HVAC (heating, ventilation and air conditioning) system regulates all of the conditions described. A *hot water* or steam heating system consisting of a boiler, piping, and radiation devices (and perhaps a pump) only controls air temperature and only during the heating season. These types of systems are common in many individual homes (residences), apartment houses, and industrial buildings.

A *warm air* system, consisting of a furnace, ducts, and air outlet registers, also controls air temperature in winter only. However, by the addition of a humidifier in the ducts, it may also control *humidity* in winter. Warm air systems are popular in residences.

Some residences have *combination* of air heating and air cooling equipment that provides control of temperature and humidity in *both* winter and summer. Some degree of control of air quality and motion is provided in air-type heating and cooling systems.

Air conditioning systems used for newer commercial and institutional buildings and luxury apartment houses usually provide *year round control* of most or all of the air conditions described. For this reason, it is becoming increasingly popular to call complete HVAC systems *environmental control systems.*

Most air conditioning systems are used for either *human comfort* or for *process control.* Air conditioning enhances our comfort. Certain ranges of air temperature, humidity, cleanliness, and motion are comfortable; others are not.

Air conditioning is also used to provide conditions that some processes require. For example, textile printing, and photographic processing facilities as well as computer rooms and medical facilities, require certain air temperature and humidity for successful operation.

COMPONENTS OF AIR CONDITIONING SYSTEMS

Heat always travels from a warmer to a cooler area. In winter, there is a continual heat loss from within a building to the outdoors. If the air in the building is to be maintained at a *comfortable temperature*, heat must be continually *supplied* to the air in the rooms. The equipment that furnishes the heat required is called a *heating system.*

In summer heat continually enters the building from the outside. In order to maintain the room air at a *comfortable temperature*, this excess heat must be continually *removed* from the room. The equipment that removes the excess heat is called a *cooling system.*

An *air conditioning system* may provide heating, cooling, or both. Its size and complexity may range from a single space heater or window unit for a small room to a huge system for a building complex. Most heating and cooling systems must have the following *basic components:*

1. A *heating source* that adds heat to a fluid (air, water, or steam).
2. A *cooling source* that removes heat from a fluid (air or water).
3. A *distribution system* (a network of ducts or piping) to carry the fluid to the rooms to be heated or cooled.
4. Equipment (fans or pumps) for moving the air or water.
5. Devices (e.g., radiation) for transferring heat between the fluid and the room.

ALL-WATER AIR CONDITIONING SYSTEMS

A typical *hydronic (all water) heating system* is shown in Fig. 52.1. Water is heated at the *heat source* (1) usually a hot water boiler. The heated water is circulated by a *pump* (2) and travels to each room through *piping* (3) and enters a *terminal unit* (4). The room air is heated by bringing it into contact with the terminal unit. Since the water loses some of its heat to the rooms, it must return to the heat source to be *reheated.*

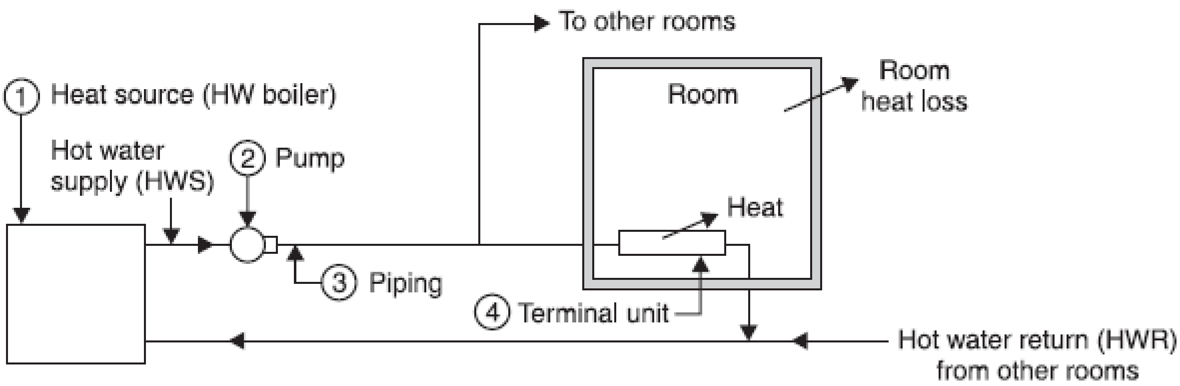


Fig. 52.1 Arrangement of basic components of a (hydronic) hot water heating system

If steam is used in a heating system, the components still work in the same manner, with the exception that a pump is not necessary to move the steam; the pressure of steam accomplishes this. However, when the steam cools at the terminal unit, it condenses into water and may require a condensate pump to return the water to the boiler.

A *hydronic cooling system* Fig. 52.2, functions in a similar manner to the hydronic heating system. Water is cooled in refrigeration equipment called a *water chiller* (1). The chilled water is circulated by a *pump* (2) and travels to each room through *piping* (3) and enters a *terminal unit* (4).

The warmer room air loses its heat to the cold water in the terminal unit. Since the water is now warmed, it must return to the water chiller to be *recooled.*

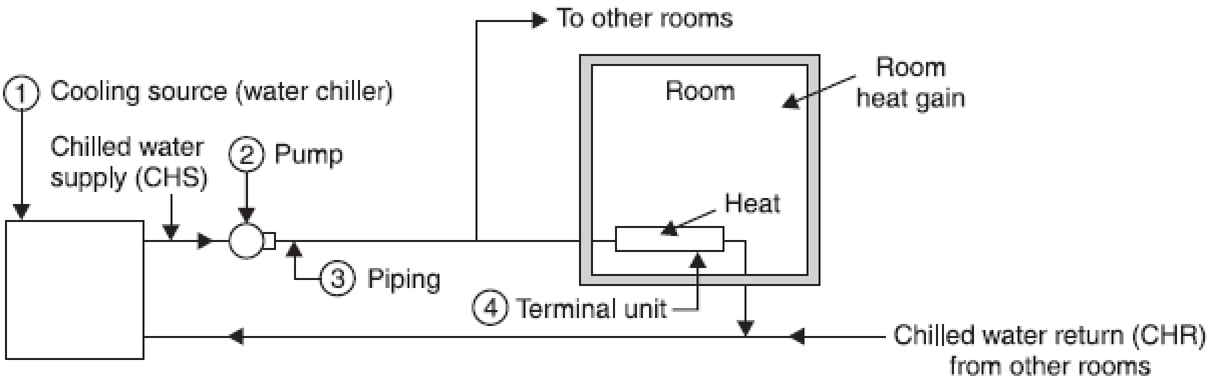


Fig. 52.2 Arrangement of basic components of (hydronic) chilled water cooling system

Hydronic systems are popular for HVAC systems that require *both* heating and cooling. This is because it is possible to use the *same piping system* for both by connecting a hot water boiler and water chiller in parallel, Fig. 52.3, using each when needed.

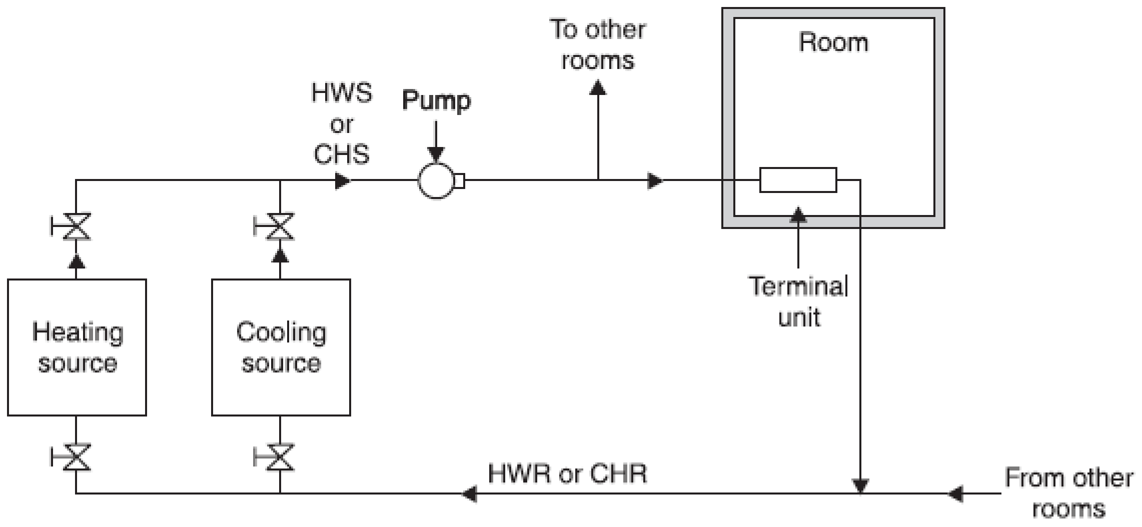


Fig. 52.3 Arrangement of basic components of a hydronic heating and cooling system

ALL-AIR AIR CONDITIONING SYSTEMS

All-air systems use *air* to heat or cool rooms. They may also have the added capability of *controlling humidity* and furnishing *outdoor ventilation*, which hydronic systems cannot do.

A typical *all-air heating and cooling system* is shown in Fig. 52.4. Air is heated at the heat *source* (1), such as a furnace. It may also be a coil circulating hot water, or steam, heated by a boiler. The heated air is circulated by a *fan* (2) and travels to each room through *supply air ducts* (3). The supply air enters the room through outlets called *air diffusers* or *registers* (4) that are designed to provide proper air distribution in the room. When the warmed supply air enters the room, the room is heated. A *humidifier* (10) may also be included to maintain a comfortable room humidity in winter.

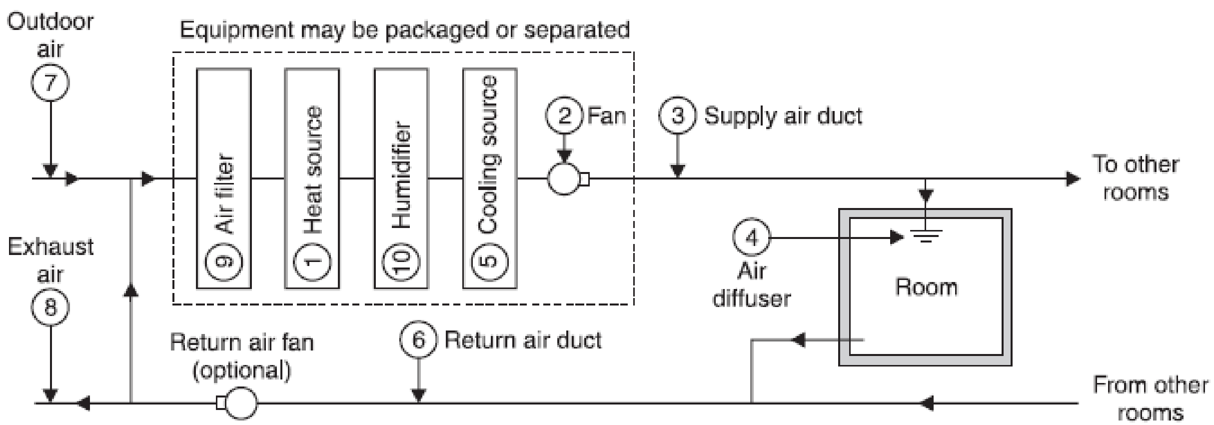


Fig. 52.4 Arrangement of basic components of an all-air heating and cooling system

In summer, air is cooled when it flows over a *cooling source* (5), usually a coil of tubing containing a fluid cooled by refrigeration equipments. When the cooled supply air enters the room, the room is cooled.

Because a room's size is fixed, the same volume of air that enters the room must also exit. This is usually accomplished by *return air ducts* (6). The air is then heated or cooled again and recirculated. An *outdoor air intake duct* (7) may be provided for introducing fresh outdoor air for increased air quality. Similarly, the same volume of air must be *exhausted* (8). Provision may be made for cleaning the air with *air filters* (9) and for *humidifying* the air (10).

REMOTE CONTROL BUTTONS

The *outdoor unit* has the other parts of the system like the compressor, air-cooled condenser, condenser fan and its own motor and is installed outside. The liquid and suction lines of the system have to be laid at site after the outdoor and indoor units are installed in position. The *distance* between the indoor and outdoor units has to be as small as possible. As this distance increases, the pressure drop in the suction and liquid lines also increases, resulting in reduction of the unit capacity.

Since the compressor (inside the outdoor unit) is remotely installed from the room to be air conditioned, the *noise level* will be appreciably lower than in the case of a room air conditioner. This is the *advantage* and the reason for opting for the split. Both room and split air conditioners and a typical remote control are shown in Fig. 52.5.

COMBINATION SYSTEMS

It is frequently desirable to combine *water and air* systems. For example, there may be instances when certain parts of a building may need *cooling* while others require *heating* simultaneously for providing comfort conditions in all areas. A typical application of this type is the air conditioning system for a big hotel. The heat removed from areas requiring cooling is transferred to the areas requiring heating. Such a system is called a *heat recovery system*, as distinguished from the *heat pump.*

COMBINATION SYSTEMS

It is frequently desirable to combine *water and air* systems. For example, there may be instances when certain parts of a building may need *cooling* while others require *heating* simultaneously for providing comfort conditions in all areas. A typical application of this type is the air conditioning system for a big hotel. The heat removed from areas requiring cooling is transferred to the areas requiring heating. Such a system is called a *heat recovery system*, as distinguished from the *heat pump.*

SPLIT AIR CONDITIONERS

In *split units* the indoor and outdoor sections of the room air conditioner are separated out into two casings or units. The *indoor unit*, Fig. 52.5, consists of the evaporator coil, evaporator blower with its own separate motor, capillary tube, control panel, air filter, supply and return grills, etc. It is installed inside the room to be conditioned. It can be ceiling suspended, wall mounted, or kept on the floor as a console unit and is generally known as *fan coil unit.*

REFRIGERATION

An *environmental control system* that includes cooling and dehumidification will require a means of removing heat from the conditioned spaces. Because heat flows from a higher to a lower temperature, a fluid with a temperature *lower than* the room design temperature must be made available, to which the *excess room heat* can be transferred.

A natural *heat sink* that is used occasionally for cooling water is atmospheric air. In climates where the humidity is extremely low, *evaporative cooling* of air may reduce both the air and water temperature low enough so that either can be used for cooling.

A *refrigeration system* extracts heat from a substance at a temperature lower than the ambient and *transfers* the extracted heat to the atmosphere at a temperature higher than the ambient. A refrigeration system is termed as :

1. *A heat recovery system* when its refrigeration effect is utilised for cooling an area or a fluid and the heat rejection is put to some beneficial use.
2. *A heat pump* when it is used for cooling during summer and heating in winter by incorporating suitable accessories for the *change over* from the cooling to heating modes and vice versa.

Food preservation, both for processing and storage, is one of the significant applications of refrigeration. Food processing calls for chilling, freezing, quick freezing, or freeze drying. Typical applications are the domestic refrigerators and home freezers, ice cream manufacturing and storage, drinking water coolers, beverage cooling, cold storages, process cooling of meat, fish, dairy products, fruits, vegetables, transport refrigeration etc. Refrigeration is very vital to the chilled and frozen-foods industry for maintaining the *cold chain* i.e. a supply of such foods from the farm to the consumer. Transport refrigeration is an important link in this chain. The conditions of temperature and relative humidity are dictated by the application for which the refrigeration system is intended.

REFRIGERANTS

Refrigerants are *heat carrying mediums* which during their cycle in the refrigeration system *absorb* heat at a low temperature level and *discard* the heat so absorbed at a higher level.

These refrigerants have been used since the 1930s because of their excellent characteristics. They have good physical properties for performance temperatures, pressure, oil mixing feature, heat transfer, specific, etc. They are non-toxic, stable, and inexpensive.

Chlorofluorocarbons (CFCs) are composed of chlorine, fluorine, and carbon atoms. Some in this group are R11, R12, and R114. *Hydrochlorofluorocarbons* (HCFCs) are composed of hydrogen, chlorine, fluorine, and carbon atoms. Some in this group are R122 and R123. *Hydrofluorocarbons* (HFCs) are composed of hydrogen, fluorine and carbon atoms. Some of these are R134a and R125.

REFRIGERATION SYSTEMS

A schematic flow diagram showing the basic components of the *vapour compression* refrigeration system is shown in Fig. 52.6. To aid in understanding, some typical temperatures for air conditioning applications are also indicated.

Refrigerant fluid *circulates* through the piping and equipment in the direction shown. There are four processes *(changes in the condition of the fluid)* that occur as it flows through the system.

Process 1–2 : At point (1), the refrigerant is in the liquid state at a relatively high pressure and high temperature. It flows to (2) through a restriction, called the *flow control device* or *expansion device.* The refrigerant loses pressure going through the restriction. The pressure at (2) is so low that a small portion of the refrigerant *flashes* (vapourises) into a gas. But in order to vapourise, it must gain heat (which it takes from the portion of the refrigerant that did not vapourise, thus cooling the mixture and resulting in low temperature at (2).

Process 2–3 : The refrigerant flows through a heat exchanger called the *evaporator.* This heat exchanger has two circuits. The refrigerant circulates in one, and in the other, the fluid to be cooled (usually air or water) flows. The fluid to be cooled is at a slightly higher temperature than the refrigerant, therefore heat is transferred from it to the refrigerant, producing the *cooling effect* desired. The refrigerant boils because of the heat it receives in the evaporator. By the time it leaves the evaporator (4), it is completely vapourised.

Process 3–4 : Leaving the evaporator, the refrigerant is *a gas at a low temperature and low pressure*. In order to be able to use it again to achieve the refrigerating effect continuously, it must be brought back to the conditions at (1)–a *liquid at a high pressure.* The first step in this process is to increase the pressure of the refrigerant gas by using a *compressor.* Compressing the gas also results in increasing its temperature.

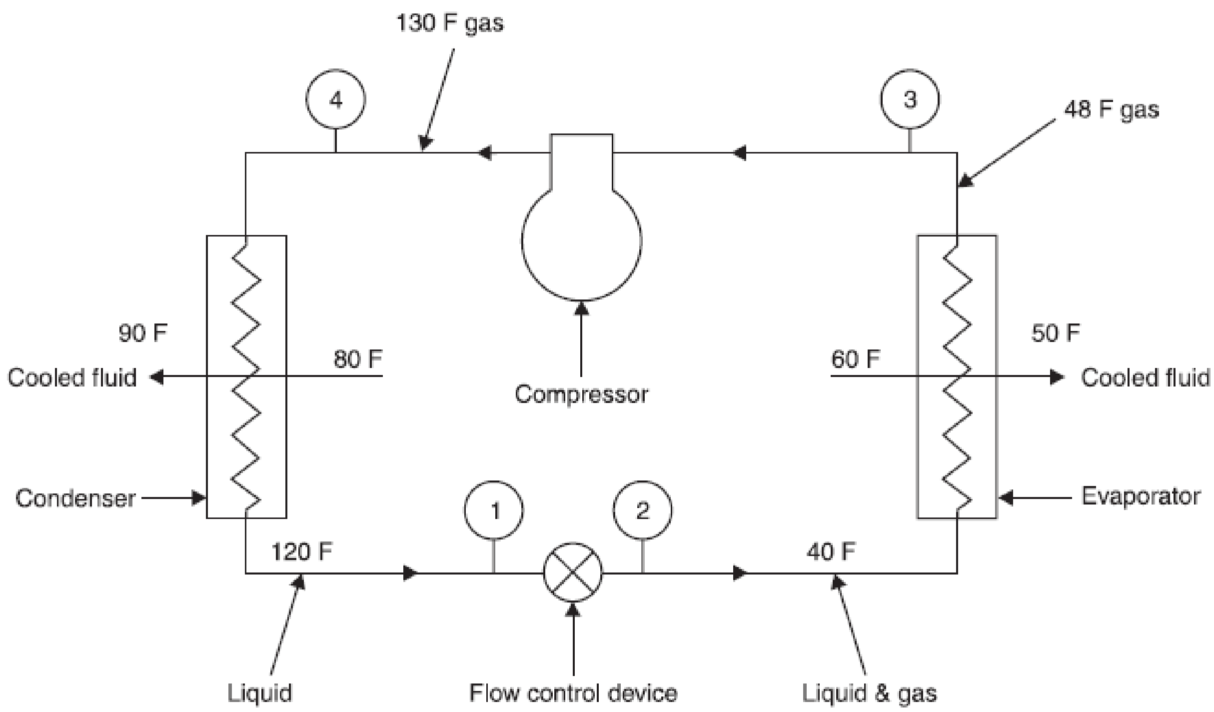


Fig. 52.6 The vapour compression refrigeration system

Process 4–1 : The refrigerant leaves the compressor as *a gas at high temperature and high pressure*. In order to change it to a liquid, heat must be removed from it. This is accomplished in a heat exchanger called the *condenser*. The refrigerant flows through one circuit in the condenser. In the other circuit, a cooling fluid flows (air or water) at a temperature lower than that of the refrigerant. Heat therefore transfers from the refrigerant to the cooling fluid, and as a result, the refrigerant condenses to a liquid (1). This is shown in Fig. 52.7.

The refrigerant has returned to its initial state and is now ready to repeat the *refrigeration cycle*. Of course the processes are actually continuous as the refrigerant circulates through the system.

The *absorption system* uses the principle that some gases will be absorbed by certain other substances. There are many *pairs of substances* that have this affinity for one another. We are all aware of how table salt absorbs water vapour from the air, thus making it difficult to pour. Yet another combination is *lithium bromide (LiBr) and water*, lithium bromide will absorb large quantities of water vapour. This pair is used in many refrigeration systems.

DOMESTIC REFRIGERATORS

The refrigerator is an essential part of almost every household for preserving food and thereby reducing wastage. The *primary function* of a refrigerator or freezer is to provide food storage space maintained at a low temperature for the preservation of food. Its essential *secondary function* is the formation of ice cubes for domestic consumption.

A storage temperature of 0° to 4°C (32° to 39°F) is satisfactory for the preservation of most of the fresh foods. For the short term storage of frozen foods, however, temperatures much below the freezing point are required. The evaporator in the domestic refrigerator, formed as a box, serves as a *freezer* for the storage of frozen food as well as for making ice cubes. It is mounted above the *food storage space*. The evaporator is held at a temperature of about –18°C (0°F) and the general storage space is cooled by natural convection. Mechanical vapour-compression cycle as well as the absorption cycle are adopted for domestic refrigerators and freezers. The mechanical *vapour-compression system* has an edge over the absorption system because of its compactness and more efficient use of electrical energy. Hence the mechanical vapour-compression system is almost universally adopted. The rear view of a single door vapour-compression refrigerator is given in Fig. 52.9.

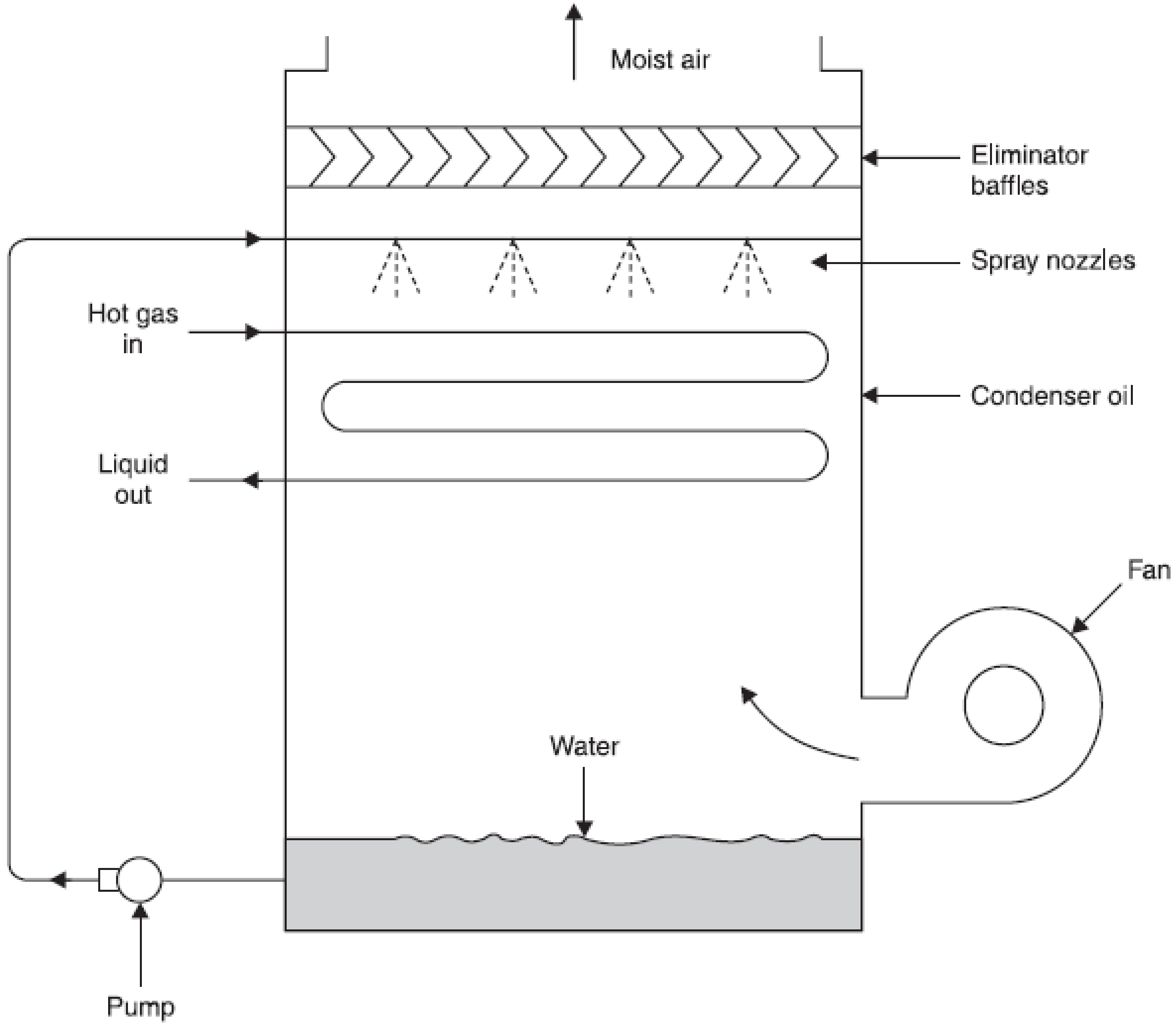


Fig. 52.7 Evaporative condenser

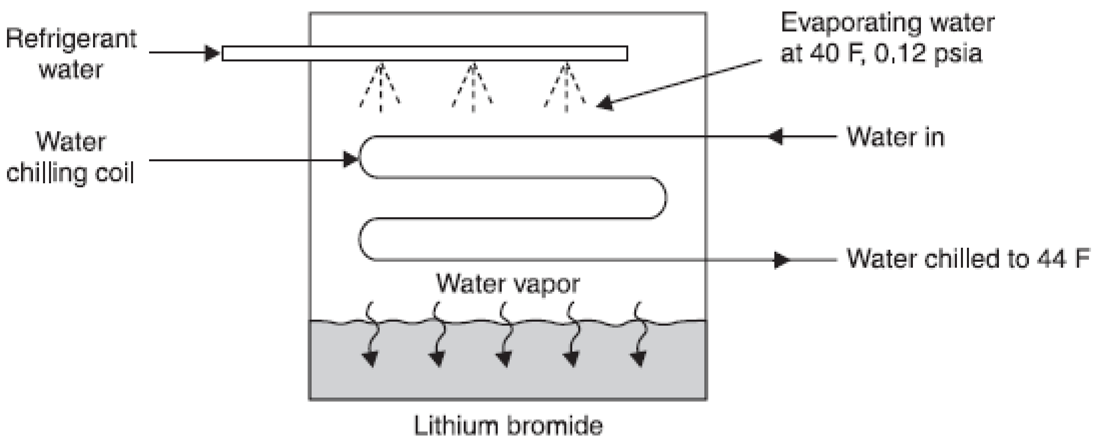


Fig. 52.8 Refrigeration by absorption

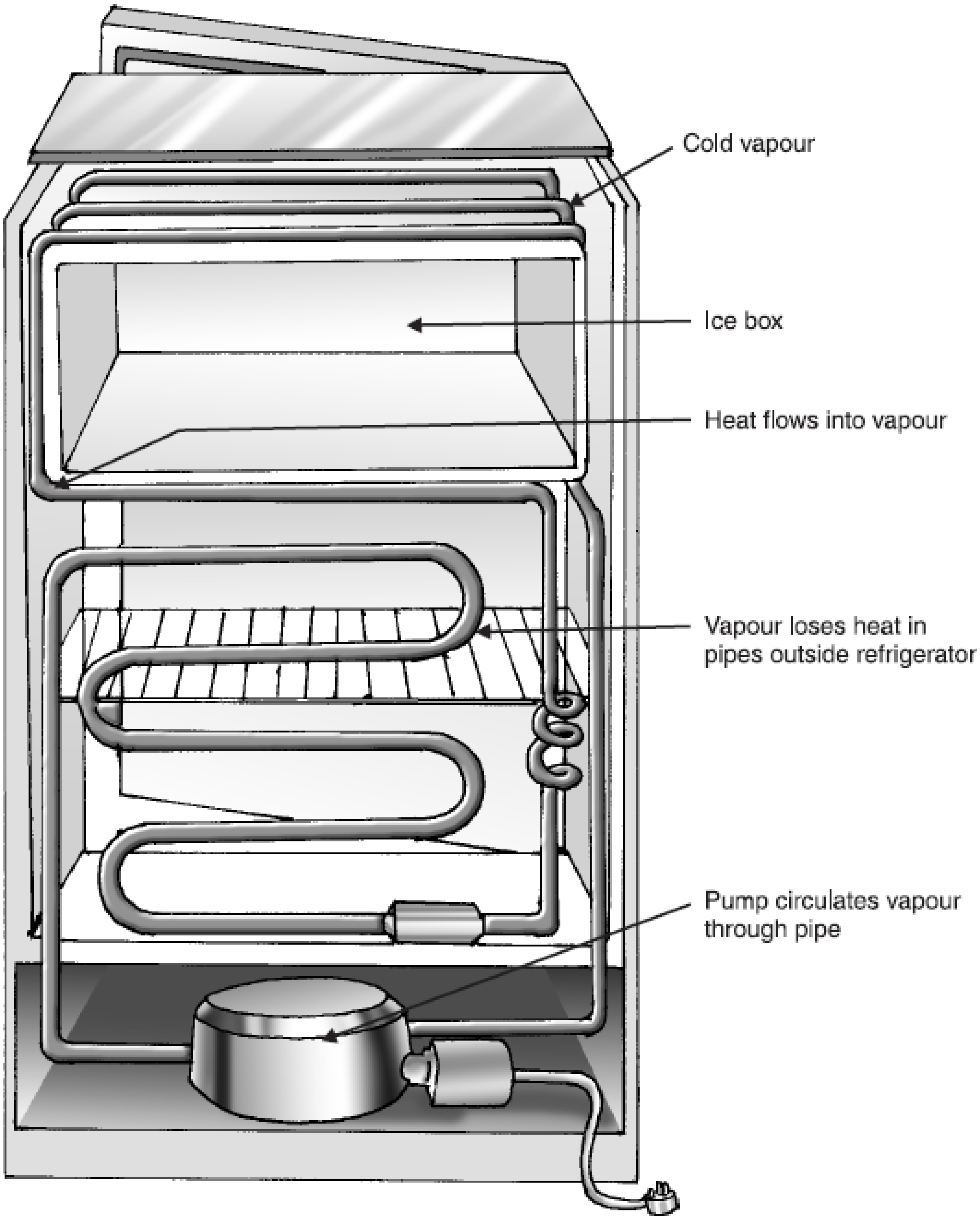


Fig. 52.9 A single-door domestic vapour-compression refrigerator (rear view)

EXERCISES

Descriptive Questions

1. What are the components of an air conditioning system?
2. Differentiate between an all-air and an all water air conditioning system.
3. What are the different types of refrigeration systems? Explain briefly.
4. Explain in detail the refrigeration cycle.
5. Explain the working of a domestic refrigerator.

6. Write short notes on:

    a. Combination systems
    b. Split air conditioners
    c. Refrigerants
    d. Evaporative condenser

Fill in the Blanks

1. Cooling means ________________ of heat.
2. Humidity is controlled by adding or removing ________________ from air.
3. Air cleanliness is controlled by either ________________ or ventilation.
4. An air conditioning system may provide heating or cooling or ________________.
5. A natural ________________ is the atmospheric air.
6. Refrigerants are ________________ mediums.
7. The refrigerant loses ________________ going through the flow control device.
8. The heat ________________ has two circuits.
9. Compressing the refrigerant gas increases its temperature and ________________.
10. The ________________ system works on the principle that some gases will be absorbed by certain other substances.
11. The vapour-compression system is ________________.
12. The ________________ is the coldest part of a refrigerator.

ANSWERS

Fill in the Blanks

1.	removal
2.	water vapour
3.	filtration
4.	both
5.	heat sink
6.	heat carrying
7.	pressure
8.	exchanger
9.	pressure
10.	absorption
11.	.economical
12.	freezer

SSS