DR.UMAYAL RAMANATHAN COLLEGE FOR WOMEN

Dr. UMAYAL RAMANATHAN COLLEGE FOR WOMEN, KARAIKUDI-03

Accredited with B+ Grade by NAAC

Affiliated to Alagappa University

(Run by Dr. Alagappa Chettiar Educational Trust)

Karaikudi – 630 003

Study Material

2020 – 2021

7BEC1E1

Antenna and Wave Propagation

Name of the Staff : **Mrs. P. Thivya**

Department : **Electronics and Communication**

Course Outcomes

Course Name	Course Code	Course Outcome	
Antenna and Wave Propagation	7BEC1E1	CO1	Learn different Wave Propagation techniques
		CO2	Study the different characteristics required for antenna to transmit and receive for wave propagation.
		CO3	Acquire knowledge on different types of Antenna needed for wave propagation.
		CO4	Describe Microwave Communication and Microwave devices.
		CO5	Study Radar concepts and Different Radar Systems used for detecting objects.

UNIT I
WAVE PROPAGATION

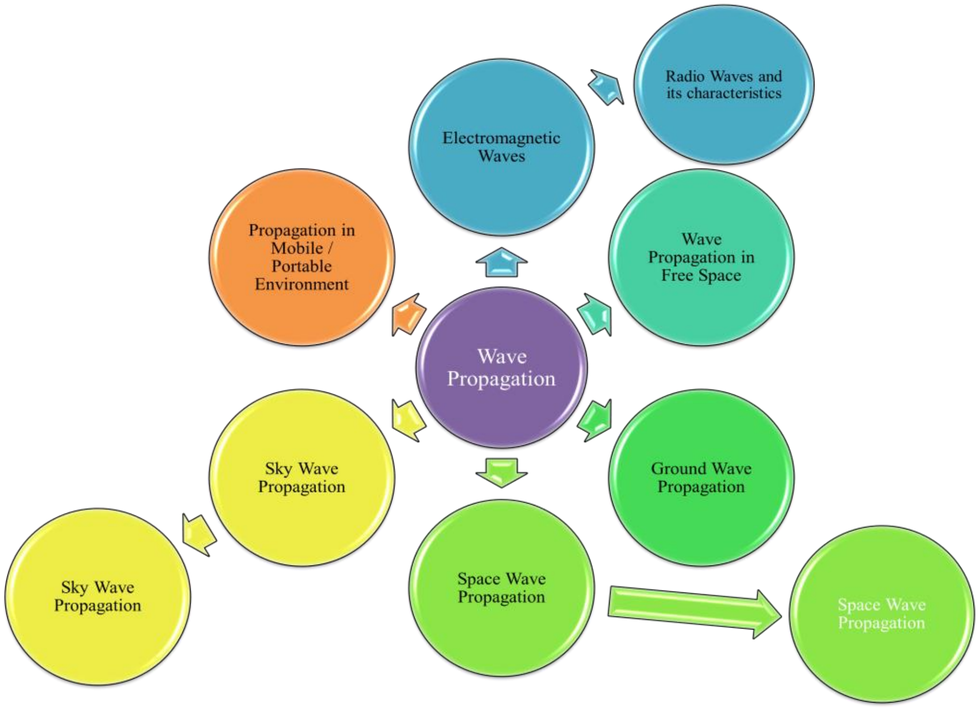
Syllabus

Electromagnetic waves – Free Space Propagation – Reflection, Refraction and Diffraction – Ground Wave Propagation – Sky wave Propagation (Ionosphere) – Space wave Propagation _ Tropospheric scatter Propagation – Line of Sight Propagation – Propagatio

Course Outcomes

Course Name	Course Code	Course Outcome	
Antenna and Wave Propagation	7BEC1E1	CO1	Learn different Wave Propagation techniques
		CO2	Study the different characteristics required for antenna to transmit and receive for wave propagation.
		CO3	Acquire knowledge on different types of Antenna needed for wave propagation.
		CO4	Describe Microwave Communication and Microwave devices.
		CO5	Study Radar concepts and Different Radar Systems used for detecting objects.

n in Mobile / Portable environment – Repeaters and Cellular system.



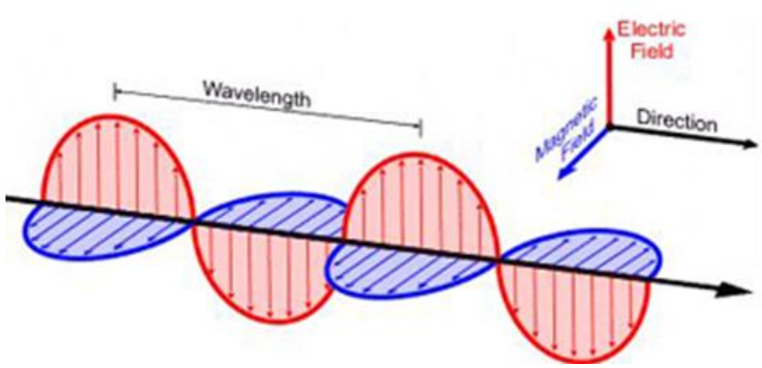
Electromagnetic (EM) Radiation

Electromagnetic (EM) radiation is a form of energy that is all around us and takes many forms, such as radio waves, microwaves, X-rays and gamma rays. Sunlight is also a form of EM energy, but visible light is only a small portion of the EM spectrum, which contains a broad range of electromagnetic wavelengths.

The study of electromagnetism deals with how electrically charged particles interact with each other and with magnetic fields.

There are four main electromagnetic interactions:

- The force of attraction or repulsion between electric charges is inversely proportional to the square of the distance between them.
- Magnetic poles come in pairs that attract and repel each other, much as electric charges do.
- An electric current in a wire produces a magnetic field whose direction depends on the direction of the current.
- A moving electric field produces a magnetic field, and vice versa.



Electromagnetic waves are formed when an electric field (shown in red arrows) couples with a magnetic field (shown in blue arrows). Magnetic and electric fields of an electromagnetic wave are perpendicular to each other and to the direction of the wave.

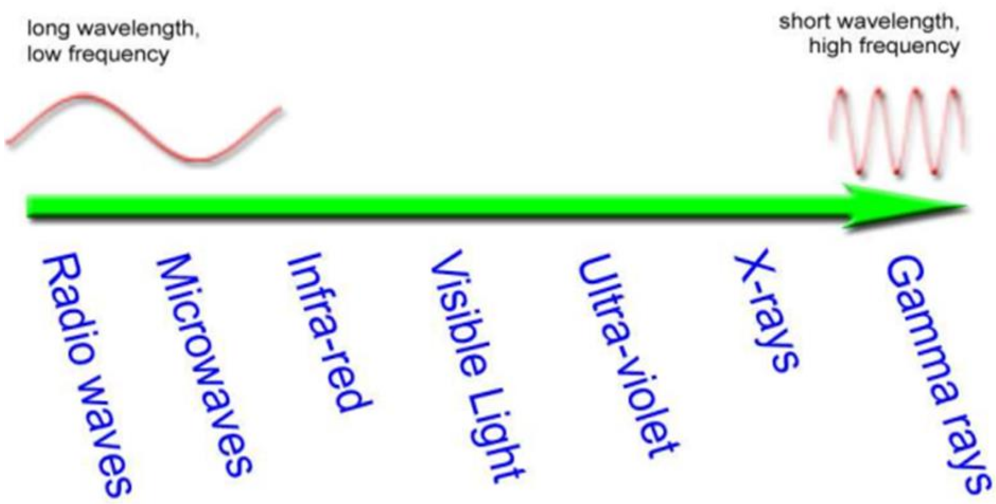
Properties of electromagnetic wave

1. All are electromagnetic waves are transverse
2. They do not require any medium to travel though.
3. They travel at speed of 3 multiple by 10 -8 in vacuum.
4. They can all be reflected and refracted.
5. They can all be emitted and absorb by matter.

The EM spectrum

EM radiation spans an enormous range of wavelengths and frequencies. This range is known as the electromagnetic spectrum. The EM spectrum is generally divided into seven regions, in order of decreasing wavelength and increasing energy and frequency. The common designations are: radio waves, microwaves, infrared (IR), visible light, ultraviolet (UV), X-rays and gamma rays. Typically, lower-energy radiation, such as radio waves, is expressed as frequency; microwaves, infrared, visible and UV light are usually expressed as wavelength; and higher-energy radiation, such as X-rays and gamma rays, is expressed in terms of energy per photon.

Chart of wave length and frequency



Region	Frequency	Wavelength	Energy
Radio waves	< 10^9	> 0.3	< 7×10^{-7}
Microwaves	10^9 - 3×10^{11}	0.001 - 0.3	7×10^{-7} - 2×10^{-4}
Infrared	3×10^{11} - 3.9×10^{14}	7.6×10^{-7} - 0.001	2×10^{-4} - 0.3
Visible	3.9×10^{14} - 7.9×10^{14}	3.8×10^{-7} - 7.6×10^{-7}	0.3 - 0.5
Ultraviolet	7.9×10^{14} - 3.4×10^{16}	8×10^{-9} - 3.8×10^{-7}	0.5 - 20
X-rays	3.4×10^{16} - 5×10^{19}	6×10^{-12} - 8×10^{-9}	20 - 3×10^{4}
Gamma rays	> 5×10^{19}	< 6×10^{-12}	> 3×10^{4}

Radio waves

Radio waves are at the lowest range of the EM spectrum, with frequencies of up to about 30 billion hertz, or 30 gigahertz (GHz), and wavelengths greater than about 10 millimeters (0.4 inches). Radio is used primarily for communications including voice, data and entertainment media.

Microwaves

Microwaves fall in the range of the EM spectrum between radio and IR. They have frequencies from about 3 GHz up to about 30 trillion hertz, or 30 terahertz (THz), and wavelengths of about 10 mm (0.4 inches) to 100 micrometers (μm), or 0.004 inches. Microwaves are used for high-bandwidth communications, radar and as a heat source for microwave ovens and industrial applications.

Infrared Waves

Infrared is in the range of the EM spectrum between microwaves and visible light. IR has frequencies from about 30 THz up to about 400 THz and wavelengths of about 100 μm (0.004 inches) to 740 nanometers (nm), or 0.00003 inches. IR light is invisible to human eyes, but we can feel it as heat if the intensity is sufficient.

Visible light

Visible light is found in the middle of the EM spectrum, between IR and UV. It has frequencies of about 400 THz to 800 THz and wavelengths of about 740 nm (0.00003 inches) to 380 nm (.000015 inches). More generally, visible light is defined as the wavelengths that are visible to most human eyes.

Ultraviolet

Ultraviolet light is in the range of the EM spectrum between visible light and X-rays. It has frequencies of about 8×10^{14} to 3×10^{16} Hz and wavelengths of about 380 nm (.000015 inches) to about 10 nm (0.0000004 inches). UV light is a component of sunlight; however, it is invisible to the human eye. It has numerous medical and industrial applications, but it can damage living tissue.

X-rays

X-rays are roughly classified into two types: soft X-rays and hard X-rays. Soft X-rays comprise the range of the EM spectrum between UV and gamma rays. Soft X-rays have frequencies of about 3×10^{16} to about 10^{18} Hz and wavelengths of about 10 nm (4×10^{-7} inches) to about 100 picometers (pm), or 4×10^{-8} inches. Hard X-rays occupy the same region of the EM spectrum as gamma rays. The only difference between them is their source: X-rays are produced by accelerating electrons, while gamma rays are produced by atomic nuclei.

Gamma-rays

Gamma-rays are in the range of the spectrum above soft X-rays. Gamma-rays have frequencies greater than about 10^{18} Hz and wavelengths of less than 100 pm (4×10^{-9} inches). Gamma radiation causes damage to living tissue, which makes it useful for killing cancer cells when applied in carefully measured doses to small regions. Uncontrolled exposure, though, is extremely dangerous to humans.

Wave Propagation in Free Space

Free space means

- Doesn't interfere with normal radiation and propagation of radio waves.
- No magnetic, electric or gravitational fields.
- No solid bodies and ionized particles.

Free space is unlikely to exist anywhere, does not exist near the earth.

Mode of Propagation

Let the average power be P_T, which is radiated spherically in all directions. At a distance of 'd', the isotropic power density in the wave be

$$P_{Di} = \frac{P_T}{4\pi d^2} \qquad \text{---(1)}$$

According to the directional characteristics of an antenna, more power is radiated in a particular direction. Thus the directivity gain will be calculated by the ratio of actual power density along the main axis to that of the power produced by an isotropic antenna at the same distance

$$G_T = \frac{P_D}{P_{Di}} \Longrightarrow P_D = G_T P_{Di}$$

$$P_D = \frac{G_T P_T}{4\pi d^2}$$

Hence, the received antenna can be fixed in a position to collect maximum power. Let P_R be the power delivered by the antenna. Then the antenna with an effective area A_{eff}, will have the P_R value be

$$P_R = P_D A_{eff} = \frac{G_T P_T}{4\pi d^2} A_{eff} \qquad \text{--- (2)}$$

For any antenna, the ratio of effective area to the maximum directivity

$\frac{A_{eff}}{G_R} = \frac{\lambda^2}{4\pi}$, λ is the wavelength of the radiated wave.

$$A_{eff} = \frac{\lambda^2 G_R}{4\pi} \text{ then,} P_R = \frac{G_T P_T}{4\pi d^2} \frac{\lambda^2 G_R}{4\pi}$$

$$\frac{P_R}{P_T} = G_T G_R \left(\frac{\lambda}{4\pi d}\right)^2$$

This is the fundamental equation for free space transmission. As we know that λ=fc, the equation can be rewritten as

$$\frac{P_R}{P_T} = G_T G_R \frac{0.57 \times 10^{-3}}{(df)^2}$$

By expressing the power ratios in decibels, we get

$$\left(\frac{P_R}{P_T}\right)_{dB} = (G_T)_{dB} + (G_R)_{dB} - (32.5 + 20\log_{10} d + 20\log_{10} f)_{dB}$$

We can write

$(32.5 + 20\log_{10} d + 20\log_{10} f)_{dB} = L$, which is known as the transmission path loss, resulting from the spreading of the wave as it propagates outward from the source (where d in kilometers and f in megahertz).

Therefore, $\left(\frac{P_R}{P_T}\right)_{dB} = (G_T)_{dB} + (G_R)_{dB} - (L)_{dB}$ ---(3)

Note: dB is converted into Watts by using the following expression

Watts = $10^{\frac{dB}{10}}$

Frequently, it is required to know the electric field strength of the wave at the receiving antenna. In terms of power density P_D and wave impedance Z_o, the electric field strength E is given as

$$E = \sqrt{Z_o P_D} \quad \text{---(4)}$$

The wave impedance Z_o can be written as

$$Z_o = \sqrt{\frac{\mu_o}{\varepsilon_o}}$$

Where $\mu_o = 4\pi \times 10^{-7}$ Hm^{-1} and $\varepsilon_o = 8.854 \times 10^{-12}$ Fm^{-1}:

Substituting these values, we get $Z_o = 120\pi\,\Omega$.

Now the field strength will become $E = \frac{\sqrt{30 P_T G_T}}{d}$ --- (5)

This is the fundamental equation that gives the field strength at the receiving antenna, for free space propagation.

Effects of the Environment

Reflection

There is much similarity between the reflection of light by a mirror and the reflection of EM waves by a conducting medium. When a wave meets a boundary, it can be reflected or transmitted. Reflection can be partial or complete. Reflection can also involve a phase flip (change of phase of 180^o). The angle of reflection is equal to the angle of incidence. The incident ray, the reflected ray and the normal at the point of incidence are in the one plane.

The reflection co-efficient 'p' is defined as the ratio of the electric intensity of the reflected wave to that of the incident wave. It is unity for a perfect conductor, and less than that for practical conductors.

If the electric vector is perpendicular to the conducting surface, no reflection will take place. For a curved conducting surface, reflection takes place and for a rough surface, reflection will be same as for the smooth surface.

Refraction

Refraction occurs when electromagnetic waves cross a boundary from one medium to another medium having a different density. A wave entering a medium at an angle changes the direction and is brought about by a change in wave velocity.

The relationship between the angle of incidence 'i' and angle of refraction 'r' is given as

$$\frac{\sin i}{\sin r} = \frac{\nu_B}{\nu_A}$$

Where ν_A is the wave velocity in medium A and ν_B is the wave velocity in medium B: As wave velocity in a dielectric medium is inversely to the square root of the dielectric constant of the medium, then

$\frac{\sin i}{\sin r} = \sqrt{\frac{\epsilon_{rA}}{\epsilon_{rB}}} = \frac{1}{\mu}$ (μ - refractive index of the medium)

For a curved boundary, refraction takes place. If the change in density of the medium is gradual, situation will be complex but refraction takes place. In the atmosphere just above the earth, there is a linear change in density, slight refraction takes place but the waves are bent down rather than travelling in a straight line.

Diffraction

Diffraction refers to the bending of waves around an edge of an object. Diffraction depends on the size of the object relative to the wavelength of the wave.

Diffraction is important in two practical situations:

1. Signals propagated by means of the space wave may be received behind tall buildings, mountains and other similar obstacles as a result of diffraction.
2. In the design of microwave antennas, diffraction plays a major role in preventing the narrow pencil of radiation which is often desired, by generating unwanted side lobes.

DIFFERENT MODES OF WAVE PROPAGATION

In an earth environment, electromagnetic waves propagate in ways that depend not only on their own properties but also on the environment itself. Propagation of radio waves takes place by different modes, the mechanism being different in each case. Based on that, it can be classified as:

1. Ground (Surface) waves
2. Space (Tropospheric) waves
3. Sky (Ionospheric) waves

Waves travel in a straight line, except where the earth and its atmosphere alter their path. Except in unusual circumstances, frequencies above HF generally travel in a straight line. They propagate by means of **space waves**. These are also called **tropospheric waves**, since they travel in troposphere. Frequencies below the HF range travel around the curvature of earth. These are called the **ground waves** or **surface waves**. All broadcast radio signals received in day time propagate by means of surface waves.

Waves in HF range, and frequencies just above or below it, are reflected by the ionized layers of the atmosphere and are called **sky waves**. To reach receivers on the opposite side of the earth, these waves must be reflected by the ground and the ionosphere several times.

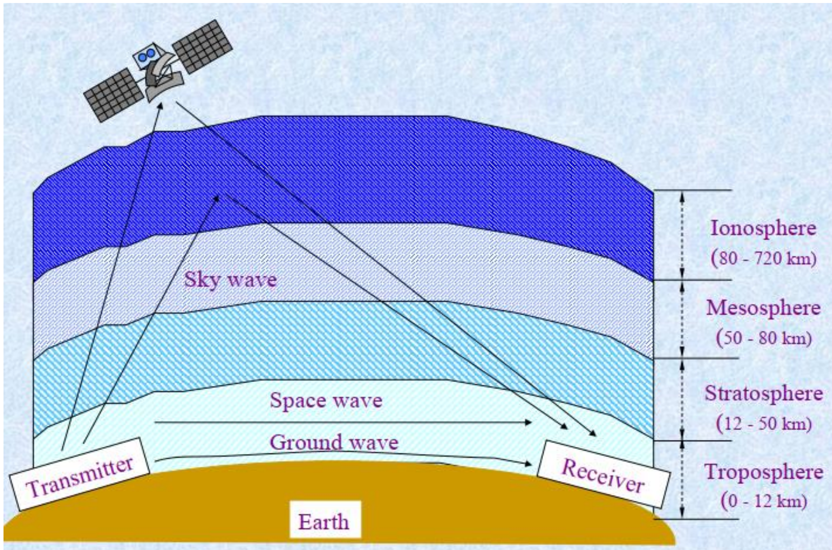
Note: Neither surface waves nor sky waves, are possible in space or on airless bodies such as moon.

Before we discuss different modes of wave propagation, let us see the allocation of frequencies for broadcasting

Classification Band	Initials	Frequency Range	Characteristics
Extremely low	ELF	< 300 Hz	
Infra low	ILF	300 Hz - 3 kHz	
Very low	VLF	3 kHz - 30 kHz	
Low	LF	30 kHz - 300 kHz	Surface/ground wave
Medium	MF	300 kHz - 3 MHz	
High	HF	3 MHz - 30 MHz	Sky wave
Very high	VHF	30 MHz - 300 MHz	Space wave
Ultra high	UHF	300 MHz - 3 GHz	
Super high	SHF	3 GHz - 30 GHz	
Extremely high	EHF	30 GHz - 300 GHz	Satellite wave
Tremendously high	THF	300 GHz - 3000 GHz	

Radio waves generally travel in four ways

- Directly from one point to another
- Follow the curvature of the earth
- Become trapped in the atmosphere and traveling longer distances
- Refracting off the ionosphere back to earth.

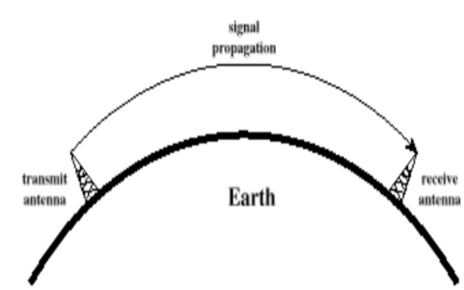


Ground Wave Propagation

Ground waves normally follows contour of the earth. They can propagate considerable distances. They cover frequencies up to 3 MHz. Example – AM radio.

At frequencies up to about 3 MHz, the most important method of propagation is by *ground waves* which are vertically polarized. They follow the curvature of the earth to propagate far beyond the horizon. Relatively high power is required.

- ➢ Medium wave (MW) propagates along the surface of the earth.
- ➢ Medium wave induces current in the ground over which it passes, and thus, loses some energy by absorption.
- ➢ Range of such coverage depends on frequency, power of the transmitter, ground conditions like salinity and conductivity of the ground or water over which the waves propagate, and the water vapour content of the air.



Received signal strength $V = \dfrac{120\pi h_r h_t I}{\lambda d}$

120π = Characteristic impedance of free space

h_t = effective height of the transmitting antenna

h_r = effective height of the receiving antenna

I = Antenna current

d = distance from the transmitting antenna

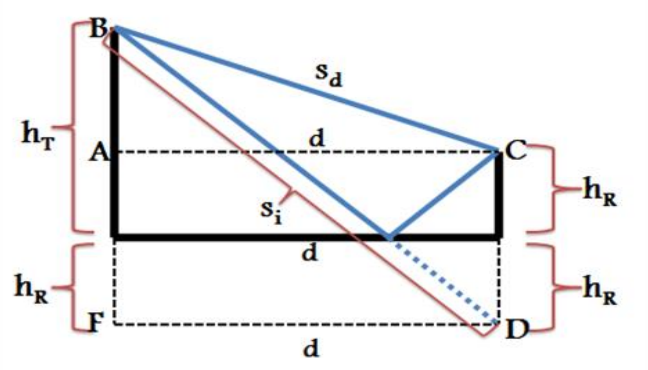
The conductivity and the permittivity of the surface play an important part in the propagation of the surface wave, as the wave will introduce both displacement and conduction currents in the surface. At low frequencies, the height of the transmitting antenna above in terms of wavelength is small, the direct wave and ground reflected wave cancel each other, leaving only the surface wave.

The medium wave (550 – 1600 kHz) broadcast service normally utilizes the surface wave. Some energy will be transmitted into the ionosphere. During day time, it is almost completely absorbed in the D-layer. During night time, a portion of ionospheric wave is returned to earth. There will be a zone, in which both the surface wave and the ionospheric wave are of the same magnitude. Then the resultant signal strength is the sum of the surface wave and the ionospheric wave.

The fluctuations in the ionosphere produce fluctuations in the phase of the reflected wave relative to the surface wave, resulting in severe fading of the combined wave. The area in which this fading occurs is known as **broadcast fading zone**.

Space (Tropospheric) Waves

They travel more or less in straight lines. As they depend on line of sight conditions, they are limited in their propagation by the curvature of the earth. Space wave can have two components: viz. direct wave and reflected wave from the surface of earth. Direct wave will be steady and strong.



The wave reaching the receiver has two components: the direct wave s_d and the ground reflected wave s_i. Consider the above figure. Normally, the reflected wave travels a greater distance than the direct wave, introducing a phase difference. Let Δs be the path difference. A phase angle of 2π corresponds to a path length of wavelength λ. Then the phase angle corresponding to Δs is

$$\phi_s = \frac{2\pi}{\lambda}\Delta s \text{ ---(6)} \qquad \text{where } \frac{2\pi}{\lambda} \text{ is the phase shift co-efficient.}$$

Let h_T is the height of the transmitting antenna, h_R is the height of the receiving antenna and d is the distance between the transmitting and receiving antenna.

From Δ^{le} FBD, $s_i^2 = (\mathrm{h_T} + \mathrm{h_R})^2 + d^2$

Similarly from Δ^{le} ABC, $s_d^2 = (\mathrm{h_T} - \mathrm{h_R})^2 + d^2$

Then $s_i^2 - s_d^2 = (\mathrm{h_T} + \mathrm{h_R})^2 - (\mathrm{h_T} - \mathrm{h_R})^2 = 4\mathrm{h_T}\mathrm{h_R}$ --- (7)

By mathematical formula, we can write $s_i^2 - s_d^2 = (\mathrm{s_i} + \mathrm{s_d}).(\mathrm{s_i} - \mathrm{s_d})$

$$s_i^2 - s_d^2 = (\mathrm{s_i} + \mathrm{s_d}).\Delta s \qquad \text{where } \Delta s = \mathrm{s_i} - \mathrm{s_d}$$

For practical purposes, mostly $\mathrm{s_i} \simeq \mathrm{s_d} \simeq$ d, then $s_i^2 - s_d^2 = 2d.\Delta s$

$$\therefore 2d.\Delta s = 4\mathrm{h_T}\mathrm{h_R} \Longrightarrow \Delta s = \frac{2\mathrm{h_T}\mathrm{h_R}}{d}$$

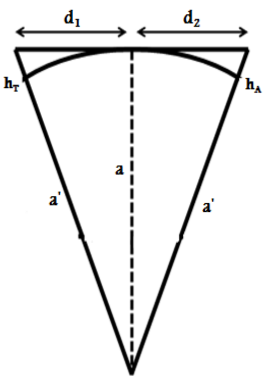
Substituting the value of Δs in eqn (6), we get

$$\phi_s = \frac{4\pi\mathrm{h_T}\mathrm{h_R}}{\lambda d} \qquad \text{--- (8)}$$

Reflection at the earth’s surface also affects the amplitude and phase of the reflected wave with respect to the direct wave. This reflection depends on the constitution of reflection surface, angle of the incident wave and the polarization of the wave (i.e., whether the wave is horizontally or vertically polarized).

Radio Horizon

The range of transmission is greater than the optical range because the effect of the earth's atmosphere is to cause a bending of the radio wave, which carries it beyond the optical horizon.



For a standard atmospheric conditions,

Fictitious radius $a' = \frac{4}{3}a$ (a - radius of the earth)

$$(a' + h_T)^2 = a'^2 + d_1^2$$

$$d_1^2 = h_T^2 + 2a'h_T$$

$$a' \gg h_T \text{ then } d_1^2 \approx 2a'h_T$$

Similarly, $d_2^2 \approx 2a'h_R$

The maximum radio range d_{max} is

$$d_{max} = d_1 + d_2 = \sqrt{2a'h_T} + \sqrt{2a'h_R} \qquad \text{--- (9)}$$

Radius of the earth, a = 3960 miles.

Expressing h_T and h_R in feet results in the following expression

$$d_{max}(miles) = \sqrt{2h_T}(ft) + \sqrt{2h_R}(ft)$$

Alternatively, in metric units

$$d_{max}(km) = \sqrt{2h_T}(m) + \sqrt{2h_R}(m)$$

Line of sight (LOS)

Radio waves normally propagate in a curved path due to refraction in the troposphere. It can be noted that not only the transmitting antenna height, but also the receiving antenna height is equally important.

$$LOS = \sqrt{2a}\left(\sqrt{h_t} + \sqrt{h_r}\right) m$$

Where

a = radius of earth = 6370 km = 6.37×10^6 m.

h_t = Transmitting antenna height in metres.

h_r = Receiving antenna height in metres.

Environment Effects

Effects of buildings: Built up area has little effect on low frequencies (few MHz). But above 30 MHz, obstruction loss and shadow loss become important. The attenuation by walls may be 2 - 5 dB at 30 MHz and increases to 10 - 40 dB at 3000 MHz.

Effects of trees and vegetation: The effect of thick vegetation is to absorb RF energy and it is particularly more dominant for vertical polarization than horizontal polarization.

Clutter losses: The loss due to natural and man-made obstruction can only be statistically evaluated and a certain allowance made in the calculations of field strength. Such losses, in general, are grouped and referred to as "Clutter losses". This loss is dependent on frequency of operation and the area surrounding the transmitter

Effective Radiated Power (ERP)

ERP is the product of intrinsic power of the transmitter and the gain of the transmitting antenna over a dipole.

ERP = Transmitter power in kW x antenna gain (In kW)

(or alternatively)

ERP = Transmitter power in dBm + antenna gain in dBm

Effective Isotropic Radiated Power (EIRP)

It is similar to ERP, except that the gain is expressed relative to an isotropic antenna.

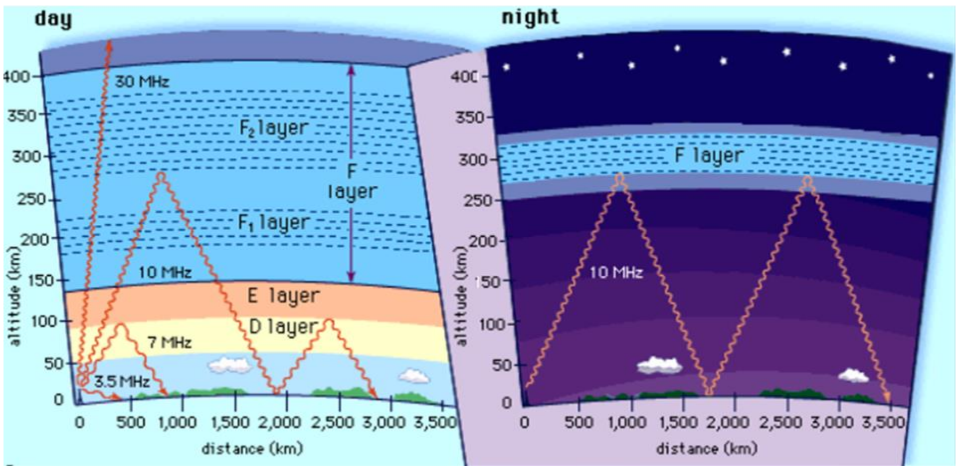
Gain of an isotrophic antenna = 1.64 times or 2.15 dB of that of a dipole.

EIRP = ERP (dBW) + 2.15 dB (In dBW) Or **EIRP = 1.64 x ERP**

Sky (Ionospheric) Waves

Short wave (SW) propagates as sky waves. Ionization of upper parts of the earth's atmosphere plays a part in the propagation of the high frequency waves. Due to the energy received from the sun, the atmospheric molecules split into positive and negative ions and remain ionized for a long period of time.

Ionospheric Layers



Ionosphere extends from 50 to 400 km and has got ionized particles. When sunrays pass through this ionosphere, due to different densities, imaginary but distinct layers are formed like D, E, F1 and F2 layers.

D Layer

It is the lowest layer of the ionosphere. Its average height is 70 km and average thickness is 10 km. Degree of ionization depends on the altitude of the sun above horizon. It disappears at night. It absorbs MF and HF waves to some extent and reflects some VLF and LF waves.

E Layer

This layer is above D-layer. Its average height is 100 km with a thickness of 25 km. It also disappears at night as the ions recombine into molecules. This is due to the absence of sun at night when radiation is no longer received. It aids MF surface wave propagation to some extent and reflects some HF waves in day time.

Es-Layer

It is a sporadic E-layer, a thin layer of very high density. Sometimes, it appears with E- layer. When Es layer occurs, it often persists during the night also. To say, it does not have an important part in long distance propagation, but sometimes permits unexpectedly good reception. Its causes are not well known.

What is sporadic E ?

Irregular scattered patches of relatively dense ionization that develop seasonally within the E region and that reflect and scatter radio frequencies up to 150 MHz. Sporadic E is a regular daytime occurrence over the equatorial regions and is common in the temperate latitudes in late spring, early summer and, to a lesser degree, in early winter. It can sometimes support reflections for distances up to 2,400 km.

F1 Layer

It exists at a height of 180 km in day time and gets combined with the F2 layer at nighttime. In day time, its thickness is about 200 km. Although some HF waves are reflected from it, most passes through it to be reflected by the F2 layer. The main effect of F1 layer is more absorption for HF waves. The absorption effect of F1 layer and any other layer is doubled because HF waves are absorbed on the way up and also on the way down.

F2 Layer

It is the most important reflecting medium for HF waves. Its approximate thickness can be up to 200 km and its height ranges from 290 to 400 km in day time. At night, it falls to about 300 km, when it combines with the F1 layer. ν Its height and ionization density vary tremendously depending upon the time of the day, the average ambient temperature and sunspot cycle.

Plasma frequency and Critical frequency

When an EM wave enters an ionized region at vertical incidence, the electric field acts as a force on the charged particles (electrons and ions), resulting in current flow. The space current is in phase opposition to the displacement current and it appears to reduce the relative permittivity (dielectric constant) of the ionized medium, given by

$$\varepsilon_r = 1 - \frac{Ne^2}{m\varepsilon_O\omega^2} \qquad \text{---- (10)}$$

Where N – Electron density (m^{-3}), e – Electron's charge magnitude (C)

M – Electron's rest mass (kg) ε_O - Permittivity of free space (8.854 x 10^{-12} Hm^{-1})

ω - Angular frequency

Let say $\omega_N^2 = \frac{Ne^2}{m\varepsilon_O}$

When $\omega = \omega_N$; ε_r is equal to zero and the term ω_N is defined as plasma angular velocity.

We can write $f_N^2 = \frac{Ne^2}{(2\pi)^2 m\varepsilon_O}$

Substituting the numerical constant values we will get $f_N = 9\sqrt{N}$

Now equation (10) can be rewritten as $\varepsilon_r = 1 - \frac{f_N^2}{f^2}$

When a wave with frequency f_N reaches the region of the electron density N, the relative permittivity is seen to be zero. The highest frequency wave that will be reflected from a ground layer will be determined by the maximum electron density of that layer and will be given by

$$f_O = 9\sqrt{N_{max}}$$

f_o is known as the critical frequency. *Critical frequency for a given layer is the highest frequency that will be returned down to earth by that layer after having been beamed straight up at it.*

Phase and Group Velocities

The phase velocity is given by $v_p = \frac{c}{\sqrt{\varepsilon_r}}$ ---(11)

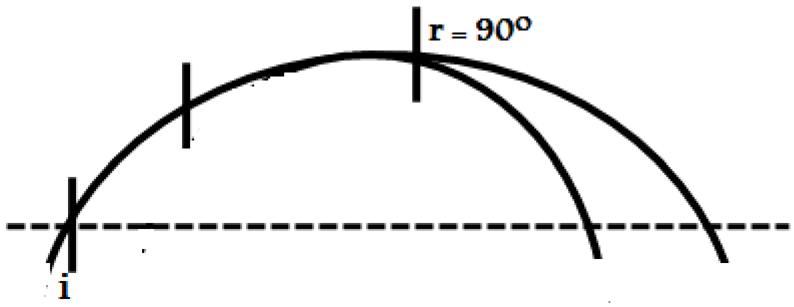
In the ionosphere, when the wave reaches a height such that ε_r is zero, V_p becomes infinite, and

$$v_p v_g = c^2 \qquad \text{--- (12)}$$

The wave front in the ionosphere is a step function that will propagate energy at the group velocity. From the equation (12), when the phase velocity is infinite, the group velocity is zero, so that the energy ceases to be propagated upward.

Secant law and Maximum Usable Frequency

When a wave enters an ionized layer at an angle of incidence i, it follows a curved path.



The phase velocity can be determined using Snell's law, $\frac{\sin i}{c} = \frac{\sin r}{v_p}$

'r' is the angle of refraction at a height where v_p occurs. At the apex of path r = 90°, then $v_p = \frac{c}{\sin i}$ -(13)

Comparing equations (11) and (13), we get $\sqrt{\varepsilon_r} = \sin i \Longrightarrow \varepsilon_r = sin^2 i$

Substituting the value of ε_r, $1 - \frac{f_N^2}{f^2} = sin^2 i$

Then $f^2 = f_N^2 sec^2 i \Longrightarrow f = f_N \sec i$

This is known as the **secant law**.

At N_{max} and f_o, the highest frequency can be used. This is known as the **maximum usable frequency (MUF)**, where MUF $= f_o \sec i$

Thus MUF depends on f_o and angle of incidence.

Keeping frequency constant, varying angle of incidence; at critical angle f becomes MUF. Below critical angle wave is reflected from lower region than N_{max} and above critical angle, the wave escapes.

Similarly keeping angle of incidence constant, varying f_O: Lower than MUF, wave is reflected from the lower point, higher than MUF, the wave escapes because of insufficient refraction.

The frequency normally used for ionospheric transmissions is known as the **optimum working frequency (OWF)** and is chosen to be 15% less than the MUF.

Virtual Height

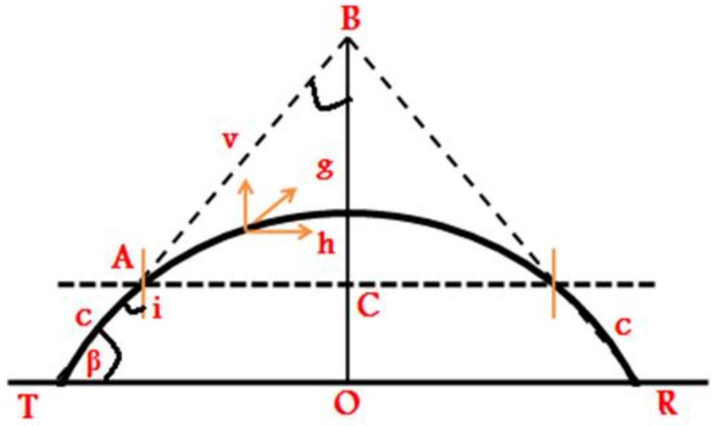
A wave travelling in a curved path has a horizontal component of group velocity given by

$$v_h = v_g \sin r$$

From equation (12), $v_g = \frac{c^2}{v_p}$ Thus $v_h = \frac{c^2}{v_p} \sin r$

From Snell's law, we can write $\frac{\sin r}{v_p} = \frac{\sin i}{c}$

So v_h becomes $v_h = c \sin i$



The time t required for the wave front to reach the highest point in the path will be

$$t = \frac{AC}{v_h}$$

where AC is the horizontal distance from point of entry to the vertical dropped from the highest point:

Now t becomes $t = \frac{AC}{c \sin i}$

In Δ^{le} ABC; AB = AC / sin i and therefore $t = \frac{AB}{c}$

From the above expression, we can say that the wave can be considered to be travelled the distance AB with velocity in time t. the virtual height h of the layer is then be OB. The virtual height has the great advantage of being easily measured and it is very useful in transmission-path calculations.

In Δ^{le} OBT; tan β=$\frac{h}{OT} \Longrightarrow OT = \frac{h}{\tan\beta}$

Similarly in Δ^{le} OBR; $OR = \frac{h}{\tan\beta}$

Then the transmission path distance $TR = \frac{2h}{\tan\beta}$

For time T, virtual height becomes $h = \frac{cT}{2}$

The virtual height is measured with the instrument namely ionosonde. The ionogram is the resultant graph that plotted between the virtual height and the frequency.
The transmission path distance d, is given as

$$d = 2a\left[\left(\frac{\pi}{2} - \beta\right) - \left(\sin^{-1}\frac{a}{a+h}\cos\beta\right)\right]$$

The skip distance is the shortest distance from a transmitter measured along the surface of the earth, at which a sky wave of fixed frequency will be returned to earth.

Fading

- Fluctuations in electron densities happen continuously in the ionosphere, giving rise to fluctuating phase differences in the various ray paths of the signal. Since the waves following these ray paths combine as phasors at receivers, fluctuations in the received signal strength, called **interference fading**.
- **Polarization fading** occurs when the ordinary and extra ordinary rays combine phasorally. The ordinary and extraordinary waves are always perpendicularly polarized with reference to each other, and because of random variations in amplitude and phase of each, when they combine; the net polarization also varies in a random manner. The receiving antenna will of course be arranged for reception of a fixed polarization wave.

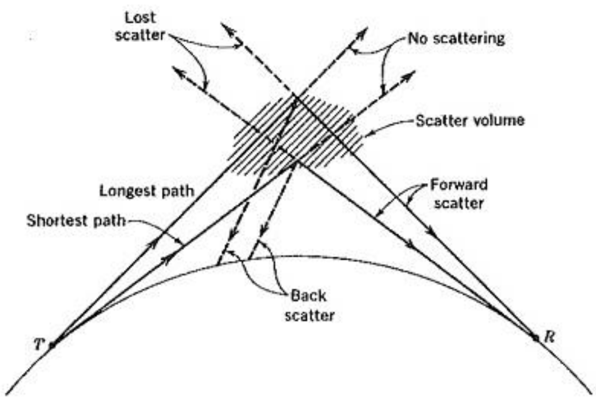
- With modulating signals, fading can affect a narrow range of the frequency spectrum independent of other parts of the spectrum and this gives to some fading called **selective fading**. Selective fading comes about essentially because the ray paths in the ionosphere will be different for different frequencies, and they will not necessarily all experience a disturbance in a given region. Selective fading limits ionospheric transmissions to narrowband signals.

Tropospheric Scatter Propagation:

Tropospheric Scatter Propagation is a means of beyond-the-horizon propagation for UHF signals. Tropospheric Scatter Propagation uses certain properties of the troposphere, the nearest portion of the atmosphere (within about 15 km of the ground).

Properties of Tropospheric Scatter Propagation:

As shown in Figure, two directional antennas are pointed so that their beams intersect midway between them, above the horizon. If one of these is UHF transmitting antenna, and the other a UHF receiving one, sufficient radio energy will be directed toward the receiving antenna to make this a useful communication system. The reasons for the scattering are not fully understood, but there are two theories. One suggests reflections from "blobs" in the atmosphere, similar to the scattering of a searchlight beam by dust particles, and the other postulates reflection from atmospheric layers. Either way, this is a permanent state of affairs, not a sporadic phenomenon.



The best frequencies, which are also the most often used, are centered on 900, 2000 and 5000 MHz. Even here the actual proportion of forward scatter to signals incident on the scatter volume is very tiny between 60 and 90 dB or one-millionth to one-billionth of the incident power. High transmitting powers are obviously needed.

Practical considerations:

Although forward scatter is subject to fading, with little signal scattered forward, it nevertheless forms a very reliable method of over-the-horizon communication. It is not affected by the abnormal phenomena that afflict HF sky-wave propagation. Accordingly, this method of propagation is often used to provide long-distance telephone and other communications links, as an alternative to microwave links or coaxial cables over rough or inaccessible terrain. Path links are typically 300 to 500 km long.

Tropospheric scatter propagation is subject to two forms of fading. The first is fast, occurring several times per minute at its worst, with maximum signal strength variations in excess of 20 dB. It is often called **Rayleigh fading** and is caused by multi path propagation.

As figure shows, scattering is from a volume, not a point, so that several paths for propagation exist within the scatter volume. The second form of fading is very much slower and is caused by variations in atmospheric conditions along the path.

It has been found in practice that the best results are obtained from tropospheric scatter propagation if antennas are elevated and then directed down toward the horizon. Also, because of the fading problems, diversity systems are invariably employed, with space diversity more common than frequency diversity. Quadruple diversity systems are generally employed, with two antennas at either end of the link (all used for transmission and reception) separated by distances somewhat in excess of 30 wavelengths.

Propagation in Mobile / Portable environment

In usual wireless transmissions, either the transmitter or the receiver or the both are in constant motion, therefore the path is also in a constant state. But in mobile environment, the situation is much cluttered because of the multiple reflections from the vehicles and the buildings. The direct signals may be blocked by tall buildings. At this condition, reflected signal is responsible for the communication to take place.

The transmission path loss equation and square law attenuation in free space is no longer applicable for mobile environments because of the multipath propagation and disturbances by obstacles. The situation is so complex and dependent on actual environment. Actually not a single equation can cover all situations. Some approximations should also be used. To understand the mobile environment, let us consider the path loss factor

$$L_{fs} = 32.44 + 20\log d + 20\log f$$

In mobile environments, the attenuation increases much more quickly with distance which the value ranges from 30 log d to 40 log d. Thus the attenuation is roughly proportional to the fourth power of the distance, because of the reflections and obstacles. The above equation had no term for antenna height, since this is irrelevant in free space where there is no ground height from.

Even for shorter distances, antenna height is an important determinant of signal strength. If the antenna height is greater, then the ground reflections will be less important. Empirical observations show that on doubling the height of the base station antenna gives about a 6dB decrease in path loss while doubling the height of the mobile or portable antenna reduces only about 3dB. From the equation, it is clear that the loss increases with frequency but actually, the effective area of an antenna decreases with the frequency.

For an ideal isotropic radiator and a receiving antenna of constant effective area, the frequency does not depend on the free-space attenuation. But in mobile environment, the situation is entirely different. The amount of reflection and diffraction of a wave from an object is a function of the object's size in wavelength.

At low frequency end, only very large objects such as the ground and buildings cause reflections. Frequency increases, wavelength decreases and such that the objects begin to reflect the waves. The increase in multipath interference may be controlled by using UHF and the other wavelength signals, which can allow the signals to propagate through windows, such as the portable wireless phone in a car or building.

When the signals penetrate building, there will be considerable loss. The loss factor depends upon the frequency, building construction, size, and location of the window and so on, roughly to be equal to 20dB at 800 MHz.

From the results obtained, it is clear that modeling of terrestrial propagation is difficult in mobile / portable environment. The present available models are the combined theory various observations at different areas such as open flat terrain, hilly terrain, suburbs and cities. Normally, a computer is used to calculate signal strengths in the coverage region. So these models are generally an approximate and 10 dB or more amount of error is common.

For an example, let us assume a mobile antenna with a height of 2m propagate wave of frequency range from 150 – 1000MHz in a dense urban mobile environment. The path loss factor is given as

$$L_P = 68.75 + 26.16 \log f - 13.82 \log h + (44.9 - 6.55 \log h) \log d$$

Where L_P – Path loss in dB

f – Frequency in MHz

h – Base station antenna height in m

d – Distance in km

From the equation, it is understandable that the increase in frequency, antenna height and distance will affect and increase the path loss.

Another important problem in the mobile / portable environment propagation is the fast fading. The signal strength tends to increase and decrease if the mobile user moves between areas of constructive and destructive interference.

Consider that the mobile receiver moves directly away from the transmitting antenna and towards a reflecting surface. Let the mobile receiver moves forward a distance of λ/4, the direct path is increase and the reflected path is reduced by the same amount, it will result in a total phase shift of 180°. When the vehicle moves another distance of λ/4, the signals are once again in phase. Thus the fade occurs if the car moves distance of λ/2. With the frequency of the signal and speed of the vehicle, the fade can be estimated as following.

The time between the fades is $T = \frac{\lambda/2}{v} = \frac{\lambda}{2v}$

$$T = \frac{c}{2fv}$$

Repeaters and Cellular Systems

Antenna height is important for line of sight communication. To achieve maximum communication range, increasing antenna height increases the distance to the radio horizon. If the distance involved is small, increased antenna height reduces multipath interference and avoids radio shadows.

Usually, in wireless communication system, the users are mobile or portable, so antenna height is not responsible for the communication. Modern wireless systems use base stations with elevated antennas. Base stations are used to provide connection between the telephone networks. These elevated antennas improve propagation.

If high transmitter power is required then it can be applied at the base station rather than at the mobile or portable environment. Normally, portable devices use only as little power as possible to reduce the size and weight of the device and especially its batteries. The simplest form of the base station has a transmitter and a receiver on the same frequency. Thus the mobile use same frequency. All portable devices have high transmitter power (in the order of 30W).

The base station antenna is placed at the maximum possible height in order to obtain wide coverage. Most of the communications take place between the mobiles and base, though mobiles can communicate directly with each other if they are close enough together. Each station can talk and listen, but not at the same time, hence the communication is called as half-duplex communication.

If each station needs to talk through the base station with greater range or need full-duplex access through the base, at least two frequencies must be used. The base station at this condition is called a

repeater, such that it transmits and receives simultaneously on at least two frequencies. Normally a single same antenna is used for both transmitting and receiving. A duplexer (high Q-filter) with resonant cavities is used to separate the transmitting and receiving frequencies. Mobiles can communicate with each other through the repeater and if the repeater is connected to PSTN (Public Switched Telephone Network) phone calls can be made through the repeater. Early mobile phone systems used 147.24 MHz and 147.3 MHz for communication and it is still in common use for fixed microwave links.

In modern cellular systems, antennas may be mounted high in order to reduce multipath and shadowing, but the range is limited by using low transmitter power. Cellular systems are complex but much more efficient in their use of spectrum. Each repeater is responsible for coverage in small cell. Let us consider a repeater is covered in all directions by the cells of hexagonal shape. But in real, the cell and repeater will not be as like this. They were in irregular shapes with some overlap. Since each transmitter operates at low power, it is possible to reuse frequencies over a short distance. Typically, a repeating pattern of either twelve or seven cells is used, and the available bandwidth is divided among these cells. The frequencies then can be reused in the next pattern.

To calculate the required number of cells in a pattern, it is necessary to make the following assumptions.

- All signals are assumed to be well above the noise level, so that co-channel interference (interference from transmitters with same frequency in other cells) is the range limiting factor.
- Assume that only the nearest cells with same frequency will cause serious problems (actually it is a true concept).
- Assume all the transmitters have equal power (in real, most cellular systems reduce transmitter power when possible to reduce interference).

Zero interference is not possible as the next cell is with the same frequency. Thus the usual assumption is that a signal-to-interference (S/I) ratio of 18 dB is at least needed.

We should know how rapidly signals drop off with distance. The greater the attenuation is better in cellular system because the interference for greater distance will be affected more than the required signal. In mobile propagation, especially in urban areas, the attenuation is proportional to the fourth power of the distance.

With this information and knowledge of symmetry, we can analyse any repeating pattern of cells t find S/I ratio. Let us consider the pattern of seven repeating cells since it is most often used in practice. In a seven cell pattern, there will be six interfering signals with equal distance from our chosen cell. All other interfering signals will be much farther away and can be ignored.

Let us consider that we are in the centre of a cell and a mobile at the edge of our cell with a distance ‘r’ from the centre. It is being interfered by the transmitters in the six nearest cells. Also assume that the interfering signals are all at the centre of their respective cell, at a distance‘d’ from us. The ratio between these distances ‘q’ is given as

$$q = \frac{d}{r} = 4.6$$

Now we can find the signal to interference S/I ratio by assuming equal transmitter power and fourth law attenuation,

$$\frac{S}{I} = \frac{d^4}{6r^4} = \frac{q^4}{6} = 74.6$$

In dB, $\frac{S}{I} = 18.7\text{dB}$

This value is only marginal when compared with practical values. Fading of the desired signal, and less attenuation may be the reason for this value. This situation can be improved by reusing the frequencies less often.

For a twelve cell repeating pattern, $\frac{S}{I} = \frac{d^4}{11r^4} = \frac{q^4}{11} = 117.8 = 20.7dB$

Another possibility which is widely employed in practice is the sectorization. Three directional antennas are located at each base station such that each antenna covers an angle of 120° and each sector uses different set of channels. The effect of this is to reduce the number of interfering signals from six to two. Now the S/I ratio is

$$\frac{S}{I} = \frac{d^4}{2r^4} = \frac{q^4}{2} = 223.8 = 23.5dB$$

Another method of calculating this improvement is

$$10\log\left({}^{6}\!/_{2}\right) = 10\log 3 = 4.8\ dB$$

Now the each cell has 3 sectors with different channels, each cell acts like three cells. This gives the equivalent in frequency reuse of 21 cell repeating pattern; so actually it is less efficient in terms of frequency reuse than the 12 cell pattern. Since cell sites are expensive, however 7 cell pattern is more economical to build.

As the number of user increases, cell sizes can be made smaller by installing more cell sites and the frequencies can be reused over closer intervals. This technique is called **cell splitting** and gives cellular system greater flexibility to adapt to changes in demand over both space and time.

In cell splitting, assuming that all cells are of equal size and that there is no overlap between cells, the number of cells required for given area is

$$N = \frac{A}{a};\ \text{N- number of cells}$$

If hexagonal cells are used, there will be no overlap at least in theory (in practical, the cells are not hexagonal). The area of hexagon (with all sides equal), is given by $a = 3.464r^2$

a – area of hexagon

R – radius of the circle inscribed in the hexagon.

Thus the number of required cells becomes $N = \frac{A}{3.464r^2}$

The number of cells required is inversely proportional to the square of the cell radius. Reducing the cell radius let say by 2, increases the number of cell and the system capacity by a factor of 4 and also the cell sites by a factor of 4. Thus when cell sizes are reduced, the available spectrum is used more efficiently, but cell-site equipment is not, since the maximum number of user per cell remains constant.

Summary

- The electromagnetic spectrum and its properties and seven classifications according to the wavelength, frequency and energy used in the wave propagation are dealt in this chapter.
- The concept of free space is explained as free space does not interfere with the wave propagation with no solid bodies and ionized particles. The expression for radiated power and transmitted power is given.
- Different modes of wave propagations such as space wave, surface wave and sky wave are explained.
- Received signal strength of ground waves is calculated.
- The maximum radio range in surface wave propagation is derived.
- Imaginary layers (D-layer, E-layer, Es-Layer, F1-layer and F2-layer) formed by the ionosphere are described.

- ✶ The expressions for plasma and critical frequency, phase and group velocity are derived. The secant law and maximum usable frequency is defined,
- ✶ The properties of tropospheric scatter propagation and propagation in mobile / portable environment are well explained.

Review Questions

Two Marks

1. What is an electromagnetic radiation?
2. What are the properties of electromagnetic wave?
3. What is an electromagnetic spectrum?
4. What are the seven different electromagnetic waves?
5. Define reflection.
6. Define refraction.
7. Define diffraction.
8. What are the different modes of propagation?
9. State ERP
10. Define EIRP.
11. What is plasma frequency?
12. Define critical frequency?
13. Define phase velocity.
14. What is group velocity?
15. Give the expression for secant law.
16. What is known as maximum usable frequency?
17. Define fading. What are the types of fading?

Five Marks

1. Explain in detail about the tropospheric scatter propagation.
2. Discuss about the wave propagation in mobile/portable environment.
3. Write a short note on line of sight propagation.
4. Write short note on repeaters and cellular system.

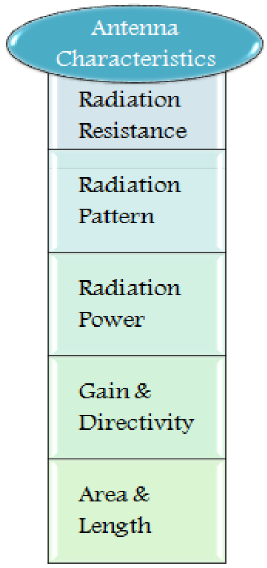
Ten Marks

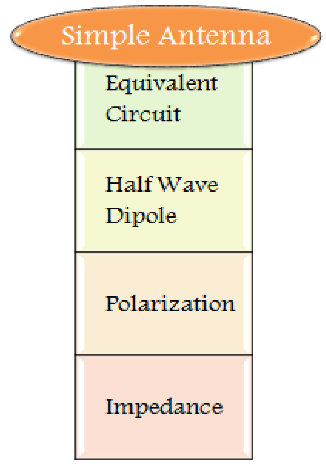
1. Explain in detail about the free space propagation.
2. Briefly explain ground wave propagation.
3. Explain space wave propagation.
4. Draw the ionospheric layer diagram and explain the sky wave propagation.

Unit II
Antenna Characteristics

Syllabus

Simple Antenna – Equivalent Circuit – The half wave dipole – Radiation Resistance – Radiation Pattern – Radiation Power – impedance – Gain and Directivity – Polarization – Area – Length of antenna.





ANTENNA - Introduction

An antenna is a metallic structure that captures and/or transmits radio electromagnetic waves. Antennas come in all shapes and sizes from little ones that can be found on your roof to watch TV to really big ones that capture signals from satellites millions of miles away. An antenna is a device that serves as interface between the electric circuit and space, and it is designed to transmit and receive electromagnetic waves. Each antenna is designed to transmit and/or to receive signals within a certain frequency range according to its size and form.

An Antenna is a transducer, which **converts electrical power into electromagnetic waves** and vice versa. An Antenna can be used either as a **transmitting antenna** or a **receiving antenna**.

- A **transmitting antenna** is one, which converts electrical signals into electromagnetic waves and radiates them.
- A **receiving antenna** is one, which converts electromagnetic waves from the received beam into electrical signals.
- In two-way communication, the same antenna can be used for both transmission and reception.

Antenna can also be termed as an **Aerial**. Plural of it is **antennae** or **antennas**. Now-a-days, antennas have undergone many changes, in accordance with their size and shape. There are many types of antennas depending upon their wide variety of applications.

At microwave frequencies apertures coupled to waveguides may be used: such antennas are naturally called **aperture antennas**. A horn antenna is an example of an aperture antenna. Antennas may be further classified as resonant antennas, in which the current distribution exists as a standing-wave pattern, and nonresonant antennas, in which current exists as a travelling wave. Again, the ordinary TV antenna is an example of a resonant antenna, usually cut to one-half wavelength, which gives it its resonant properties. Nonresonant antennas are mainly used for very short wave communications links.

Need of Antenna

In the field of communication systems, whenever the need for wireless communication arises, there occurs the necessity of an antenna. Antenna has the capability of sending or receiving the electromagnetic waves for the sake of communication, where you cannot expect to lay down a wiring system.

Main characteristics of antennas

Directionality: an antenna may be non-directional (isotropic), transmitting uniformly in all directions, or directional –transmitting preferentially in one or more directions at a higher power.

Radiation curve of an antenna: a graphic three dimensional representation of the radiation intensity of the antenna, representing the distribution of the electric field and/or the radiation energy in the space surrounding it.

Enhancement (amplification) of the antenna: The ratio between the radiation power of an antenna in a certain direction, and the radiation power of an isotropic (omnidirectional) antenna under identical conditions of transmission (distance and input power of the antenna).

Polarization: direction of the electric field radiating from the antenna in areas sufficiently distant from the antenna (a distance of a number of wavelengths). A radio wave may be non-polarized or polarized (linear or circular/elliptical polarization).

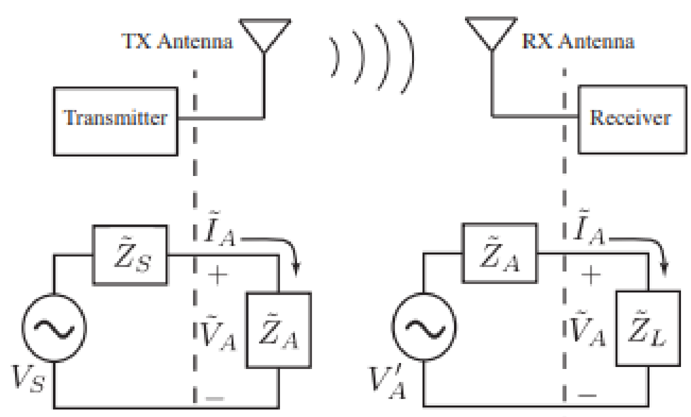
Efficiency: The ratio between the total power radiated by the antenna and the net electric power fed to the antenna.

Band width: The width of a range of frequencies at which the antenna transmits its maximal radiation and intensity.

Antenna array: A systematic deployment of antennas that operate together. The individual antennas in an array are usually of the same type, and are situated in close proximity and at a fixed distance from one another. An antenna array enables increasing the directionality and control of the main radiation beams and the lateral beams.

Equivalent Circuit

In a radio communication link, the transmitting antenna is coupled to the receiving antenna through the electromagnetic wave. An overall sketch of two equivalent time-harmonic circuits used for antennas – one for transmit antennas and the other for receiver antennas is given in the figure below. The antenna coupling system can be represented as a four terminal network. For a transmitting and receiving -mode, the antenna is modeled as complex impedance Z_A that is connected to a sinusoidal voltage source of amplitude V_S, source impedance Z_S and load resistance Z_L. The phasors V_A and I_A are the amplitudes and phases of the voltage and current, respectively, at the terminals of the antenna.



Based on these values, we can calculate the average power delivered into the antenna, P_T, by the source:

$$P_T = \frac{1}{2}\Re\left\{\tilde{V}_A \tilde{I}_A\right\} = \frac{V_S^2 R_A}{2\left|\tilde{Z}_S+\tilde{Z}_A\right|^2} \qquad \begin{array}{l} \tilde{Z}_S = R_S + jX_S \\ \tilde{Z}_A = R_A + jX_A \end{array}$$

The power P_T is the transmit power of the system and includes mismatch losses between the antenna and the source. Not all of this power will necessarily be radiated by the transmit antenna; some of the power may be absorbed, particularly if the antenna has a low radiation efficiency. However, a well-designed antenna will radiate most of P_T.

The maximum possible power that the source can deliver occurs when the antenna impedance Z_A is conjugate matched with the source impedance Z_S. Under these conditions, the maximum possible power delivered by the source, P_S, is given by

$$\text{Conjugate Match } \tilde{Z}_A = \tilde{Z}_S^*: \quad P_S = \frac{V_S^2}{8R_S}$$

In terms of the total available source power, P_S, we can express the transmitted power as

$$P_T = \underbrace{\frac{4R_S R_A}{\left|\tilde{Z}_S + \tilde{Z}_A\right|^2}}_{\text{mismatch losses}} P_S$$

Power can also be tracked at a receiver antenna using the equivalent circuit on the right-hand side of Figure. The real average power, P_L, delivered to the load impedance representing the receiver hardware is given by

$$P_L = \frac{1}{2}\Re\left\{\tilde{V}_A \tilde{I}_A\right\} = \frac{V_A'^2 R_L}{2\left|\tilde{Z}_A+\tilde{Z}_L\right|^2} \qquad \begin{array}{l} \tilde{Z}_L = R_L + jX_L \\ \tilde{Z}_A = R_A + jX_A \end{array}$$

If a transmit antenna with impedance Z_A is used for receiving purposes, it will have the same impedance value Z_A when its role is reversed. This convenient property is a direct result of reciprocity in electromagnetism. The maximum available received power, P_R, from the antenna can only be delivered to the load under conjugate-matched conditions:

$$\text{Conjugate Match } \tilde{Z}_L = \tilde{Z}_A^*: \quad P_R = \frac{V_A'^2}{8R_A}$$

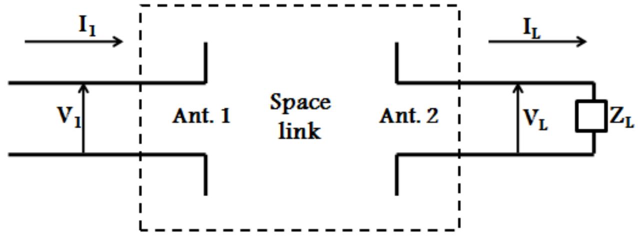
This maximum received power, P_R, is commonly available as a specification of a communications link or as the product of a link budget calculation. The following equation gives us a more convenient relationship for V_A in terms of received power:

$$V_A' = 2\sqrt{2R_A P_R}$$

Using this result, we can now calculate average power delivered to the receiver load, P_L, in terms of received power, P_R:

$$P_L = \underbrace{\frac{4R_L R_A}{\left|\tilde{Z}_A + \tilde{Z}_L\right|^2}}_{\text{mismatch losses}} P_R$$

A network theorem known as the reciprocity theorem can also be applied to the antenna system shown in fig. *This theorem states that if an emf E is applied to the terminals of antenna 1 gives rise to a terminal current I at antenna 2, then applying E to the terminals of antenna 2 will give rise to the* *current I at the terminals of antenna 1*. Now it is known that all practical antennas are directive i.e., they radiate better in some directions than others and receive better from some directions than others. A consequence of reciprocity theorem is that the directive pattern for a given antenna will be the same for both the transmitting and the receiving modes of operating.



The antenna impedance Z_A is a complex quantity: $Z_A = R_A + jX_A$

The reactive part X_A results from the reactive fields surrounding the antenna. As with any reactance, energy is stored in these fields and returned to source. Wherever possible, the reactance will be tuned out, so that the antenna presents a purely resistive load to the transmission line. The resistive part R_A is given by

$$R_A = R_{loss} + R_{rad}$$

The resistance R_{rad} is a fictitious resistance termed as **radiation resistance**, which if it carried the same rms terminal current as the antenna, on transmission would dissipate the same amount of power as was radiated. A certain amount of power will be dissipated in the antenna as heat, and the power dissipated in R_{loss}, when carrying the same current as R_{rad}, gives the power lost in this way. The resistance R_{loss}, therefore represents the losses in the antenna.

The concepts of loss resistance and radiation resistance are most useful with wire antennas, for which the terminal currents are easily identified and the loss resistance is mainly the resistance of the antenna wire. For this type of antenna, let I be the rms terminal current. Then the total power supplied to the antenna is I^2R_A and the power radiated is R_{rad}. The antenna efficiency is therefore,

$$\eta_A = \frac{I^2 R_{rad}}{I^2 R_A} = \frac{R_{rad}}{R_A}$$

In the receiving mode, the **efficiency** is defined as the ratio of power delivered to the matched load from the actual antenna to the power delivered to the matched load from the antenna with R_{loss} assumed equal to zero. Applying maximum power transfer theorem to the receiving antenna circuit of fig 1(c) for the real antenna the maximum power is $\frac{V_S^2}{4R_A}$, for the lossless antenna it $\frac{V_S^2}{4R_{rad}}$. Thus the receiving efficiency is also given by $\eta_A = \frac{R_{rad}}{R_A}$

Antenna matching to the feeder line is important for eliminating reflected waves and obtaining maximum power transfer. In general, the matching network has to provide both reflection-less matching and matching for maximum power transfer. A matching network designed to meet one of these conditions automatically satisfies the other condition. Thus, as shown in fig (2), the impedance seen looking into the network at the feeder side is Z_o and at the antenna side is $Z_A^* = R_A - jX_A$ is the complex conjugate of Z_A. This is required for maximum power transfer.

In effect, the matching network tunes the antenna to resonance and then transforms the resistive part to Z_O and can be represented by the general arrangement. Working from antenna to transmission line, the reactance $-jX_A$ tunes out the $+jX_A$ component of antenna impedance, and the transformer transforms the remaining R_A component to Z_O. Now, assuming that the transmission line is matched at its far end, the matching network sees impedance Z_O, and working from line to antenna, the transformer transforms this Z_O back into R_A. This R_A is in series with the $-jX_A$ element, and therefore the output impedance is $R_A - jX_A$ orZ_A^*, as required for maximum power transfer to Z_A.

Consider now the effect of mismatch at the antenna end. Under transmitting conditions, the transmitter and line appear as an emf source of internal resistance Z_O feeding a load Z_A.

The current flowing in this circuit is $\frac{V_O^2}{|Z_O+Z_A|^2}$ and the power delivered to Z_A is $\frac{R_A V_O^2}{|Z_O+Z_A|^2}$. The power delivered under matched conditions would be $\frac{V_O^2}{4Z_O}$, and therefore the matching efficiency can be written as

$$\eta_\Gamma = \frac{R_A V_O^2}{|Z_O + Z_A|^2} \frac{4Z_O}{V_O^2} = \frac{4R_A Z_O}{|Z_O + Z_A|^2}$$

Under receiving conditions, the power delivered to Z_O is $\frac{Z_O V_A^2}{|Z_O+Z_A|^2}$, and the power would have been delivered under matched conditions is $\frac{V_A^2}{4Z_A}$. Hence the matching efficiency in the receiving case is also $\eta_\Gamma = \frac{Z_O V_A^2}{|Z_O+Z_A|^2} \frac{4Z_A}{V_A^2} = \frac{4R_A Z_O}{|Z_O+Z_A|^2}$

Thus the matching efficiency is the same for both transmitting and receiving conditions. Using the relationships $Z_A = Z_O \frac{1+\Gamma_A}{1-\Gamma_A}$ and $R_A = \frac{1}{2}(Z_A + Z_A^*)$, the matching efficiency is given by $\eta_\Gamma = 1 - |\Gamma_A|^2$

Isotropic Radiator

An isotropic radiator is one that radiates equally in all directions. (Isotropic means equally in all direction). A star is a perfect example of an isotropic radiator of electromagnetic energy. All real antennas radiate better in some directions than others and cannot be isotropic. The need of the isotropic radiator concept is to improve real antennas to a standard by comparing the concepts. As it is a hypothetical radiator, let us also assume it to be a lossless antenna. Therefore, efficiency becomes unity; which is also the power radiated.

1. P_S is the power input to a lossless isotropic radiator. Consider the antenna at the centre of the spherical surface. Sphere has a solid angle of 4π steradians at its centre. The power per unit solid angle is $\boldsymbol{P_i = \frac{P_S}{4\pi}}$ W/sr.
2. Power density for the lossless isotropic radiator is given by $\boldsymbol{P_{Di} = \frac{P_S}{4\pi d^2}}$, where $\boldsymbol{4\pi d^2}$ is the surface area of the sphere of radius ‘d’.

$$\boldsymbol{P_{Di} = \frac{P_i}{d^2}}$$

These two quantities are used as a standard values to design a real antenna.

Power Gain of an Antenna

The power per unit solid angle will vary depending on the direction in which it is measured and therefore it will be the function of the angular coordinates θ and φ as P(θ,φ).

The power gain of the antenna is defined as the ratio of P(θ,φ) to the power per unit solid angle radiate by a lossless isotropic radiator. The gain function G(θ,φ) is given by

$$\boldsymbol{G(\theta,\varphi) = \frac{P(\theta,\varphi)}{P_i} = \frac{4\pi P(\theta,\varphi)}{P_S}}$$

The gain function is the important antenna characteristics which can be measured, sometimes calculated. For most antennas, the gain function shows a well defined maximum (G_M) and the radiation pattern of the antenna is $\boldsymbol{g(\theta,\varphi)} = \frac{\boldsymbol{G(\theta,\varphi)}}{\boldsymbol{G_M}}$.

The radiation pattern is simply the gain function normalized to its maximum value.

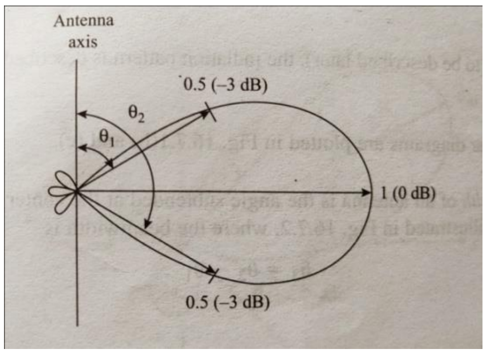
The maximum value G_M is the gain of the antenna which represents that the antenna concentrates or focuses the power in the maximum direction. Actually, it doesn't increase the total power radiated.

Another important term is the directive gain of the antenna. This is the ratio of P(θ,φ) to the average power per unit solid angle radiated by the actual antenna and is denoted by D(θ,φ). The average power per unit solid angle is given by $\boldsymbol{\eta_A.\frac{P_S}{4\pi}}$ or $\boldsymbol{\eta_A.P_i}$

$$\boldsymbol{D(\theta,\varphi)} = \frac{\boldsymbol{P(\theta,\varphi)}}{\boldsymbol{\eta_A P_i}} = \frac{\boldsymbol{G(\theta,\varphi)}}{\boldsymbol{\eta_A}}$$

Similarly, the maximum value of D(θ,φ) is termed as the directivity or directive gain given by $\boldsymbol{D_M} = \frac{\boldsymbol{G_M}}{\boldsymbol{\eta_A}}$

When the gain function is plotted, the result will be a three dimensional plot. From the plot, the length of the line from the origin to any point on the surface gives the gain in the direction of the point. The gain G_1 [$G(\theta_1,\varphi_1)$] in (θ_1,φ_1) direction is the maximum gain G_M.. G_2, G_3 are the minor lobes.



In practice, two dimensional plots are often used, one for the equatorial plane and one for the meridian plane. In the equatorial plane, the radiation pattern g(θ,φ) is denoted by g(φ), since θ is constant and in the meridian plane, it is denoted by g(θ), since φ is constant.

Directivity

"The ratio of maximum radiation intensity of the subject antenna to the radiation intensity of an isotropic or reference antenna, radiating the same total power is called the **directivity**."

An Antenna radiates power, but the direction in which it radiates matters much. The antenna, whose performance is being observed, is termed as **subject antenna**. Its **radiation intensity** is focused in a particular direction, while it is transmitting or receiving. Hence, the antenna is said to have its **directivity** in that particular direction.

- The ratio of radiation intensity in a given direction from an antenna to the radiation intensity averaged over all directions is termed as directivity.
- If that particular direction is not specified, then the direction, in which maximum intensity is observed, can be taken as the directivity of that antenna.
- The directivity of a non-isotropic antenna is equal to the ratio of the radiation intensity in a given direction to the radiation intensity of the isotropic source.

The radiated power is a function of the angular position and the radial distance from the circuit. Hence, it is expressed by considering both the terms **θ** and **Ø**.

$$Directivity = \frac{Maximum\ radiation\ intensity\ of\ subject\ antenna}{Radiation\ intensity\ of\ an\ isotropic\ antenna}$$

$$D = \frac{\phi(\theta, \phi)_{max}(from\ subject\ antenna)}{\phi_0(from\ an\ isotropic\ antenna)}$$

Where

- $\phi(\theta,\phi)_{max}$ is the maximum radiation intensity of subject antenna.
- ϕ_0 is the radiation intensity of an isotropic antenna (antenna with zero losses).

Gain

"**Gain** of an antenna is the ratio of the radiation intensity in a given direction to the radiation intensity that would be obtained if the power accepted by the antenna were radiated isotropically."

Simply, gain of an antenna takes the directivity of antenna into account along with its effective performance. If the power accepted by the antenna was radiated isotropically (that means in all directions), then the radiation intensity we get can be taken as a referential.

- The term **antenna gain** describes how much power is transmitted in the direction of peak radiation to that of an isotropic source.
- **Gain** is usually measured in **dB**.
- Unlike directivity, antenna gain takes the losses that occur also into account and hence focuses on the efficiency.

$$G = \eta_e D$$

Where

- **G** is gain of the antenna.
- η_e is the antenna's efficiency.
- **D** is the directivity of the antenna.

The unit of gain is **decibels** or simply **dB**.

Radiation is the term used to represent the emission or reception of wave front at the antenna, specifying its strength. In any illustration, the sketch drawn to represent the radiation of an antenna is its **radiation pattern**. One can simply understand the function and directivity of an antenna by having a look at its radiation pattern.

The power when radiated from the antenna has its effect in the near and far field regions.

- Graphically, radiation can be plotted as a function of **angular position** and **radial distance** from the antenna.
- This is a mathematical function of radiation properties of the antenna represented as a function of spherical co-ordinates, E (θ, Ø) and H (θ, Ø).

Effective Area of an Antenna

Effective area is an area where the receiving antenna collects electromagnetic energy from the incident wave, such as a solar panel collects energy from sunlight.

Let us assume that the antenna is in the far-field zone, the wave incident on it will be a plane transverse electromagnetic wave with power density P_D. Let the receiving antenna be at the centre of a spherical coordinate system and the incoming wave be denoted by angular coordinates (θ,φ). The power delivered to the matched load at the receiver end will be the function of direction, so we can take the effective area also be the function of direction i.e., A=A(θ,φ). Thus if P_R is the power delivered to a matched load,

$$P_R = P_D\, A(\theta,\varphi)$$

This is the defining equation for effective area. This effective area will leave maximum value A_{eff} called the effective area of an antenna.

As a result of reciprocity theorem, the effective area normalized to its maximum value can be written as $\frac{A(\theta,\varphi)}{A_{eff}} = g(\theta, \varphi)$

Such that A_{eff} is proportional to G_M. again from reciprocity theorem, the proportionality constant for all antennas is $\frac{\lambda^2}{4\pi}$.

Then $\frac{A_{eff}}{G_M} = \frac{\lambda^2}{4\pi}$

If the gain of an antenna is measured under transmission conditions as G_T, the effective area under receiving conditions can be written as $A_{eff} = \frac{\lambda^2 G_T}{4\pi}$

Often in theoretical calculations, the directivity D_T is used instead of G_T and this will give a higher value of effective area.

The factors that can reduce effective area are

- The mismatch factor $\eta_\Gamma = 1 - |\Gamma_A|^2$, where Γ is the voltage reflection coefficient which is given by $\Gamma = \frac{V_r}{V_i}$ (V_r is the reflection voltage and V_i is the incident voltage).

  If the antenna is matched to the line, no reduction will occur.
- Polarization misalignment factor: already we have defined effective area as a function of angular coordinates (θ, φ). Thus any loss from polarization misalignment also affects the effective area. The reduction in effective area as a result of polarization misalignment is given as $(plf)^2$ where plf is the polarization loss factor which is given by $plf = cos\psi$

Hertzian Dipole

Hertzian dipole is a short linear antenna assumed to carry uniform current along its length when radiating, such an antenna is not possible in practical, but can be assumed to be made of a number of Hertzian antenna connected in series.

An approximation to a Hertzian dipole can be achieved by capacitively loading the ends of a short center fed dipole. The capacitive end allows nearly an uniform charging current along the wire. Let the length δl be shorter than the wavelength ($\delta l \gg \lambda$) over the entire dipole, the current is assumed to be $i = I_o \sin \omega t$

The electric field in the far-field zone is directly proportional to the component of current that is parallel to the electric field is $I_o \sin \theta$. From the physics of radiation, it is also found the instantaneous electric field is proportional to the rate of change of current and inversely proportional to the distance d, then

$$e_\theta = \frac{60\pi \delta l I_o \sin \theta}{\lambda d} \cos \omega \left(t - \frac{d}{c}\right)$$

When T=d/c, instantaneous electric field value becomes the maximum electric field, which is given by

$$E_o = \frac{60\pi \delta l I_o \sin \theta}{\lambda d}$$

Now the rms current I is replaced by maximum current I_O, to give rms field strength E

$$E = \frac{60\pi \delta l I \sin \theta}{\lambda d}$$

The power density is given by $P_D = \frac{E^2}{120\pi}$

$$P_D = \frac{1}{120\pi} \frac{60\pi\delta l I \sin\theta}{\lambda d} . \frac{60\pi\delta l I \sin\theta}{\lambda d} = 30\pi\left(\frac{\delta l I}{\lambda d}\right)^2 sin^2\theta$$

$$P_D = P_{DM} sin^2\theta$$

Where $P_{DM} = 30\pi\left(\frac{\delta l I}{\lambda d}\right)^2$

The total power is obtained by $P_T = \int_0^\pi P_D\, dA$ and $dA = 2\pi d^2 \sin\theta d\theta$

$$P_T = \int_0^\pi 30\pi\left(\frac{\delta l I}{\lambda d}\right)^2 sin^2\theta . 2\pi d^2 \sin\theta d\theta$$

$$P_T = 30\pi\left(\frac{\delta l I}{\lambda d}\right)^2 . 2\pi d^2 \int_0^\pi sin^3\theta d\theta$$

$$\boxed{\begin{gathered}\int sin^n\theta d\theta = \left[\frac{-sin^{n-1}\theta\cos\theta}{n}\right] + \frac{n-1}{n}\int \sin\theta\, d\theta \\ \int sin^3\theta d\theta = \left[\frac{-sin^2\theta\cos\theta}{3}\right] + \frac{2}{3}\int \sin\theta\, d\theta\end{gathered}}$$

$$\int_0^\pi sin^3\theta d\theta = \left[\frac{-sin^2\theta\cos\theta}{3}\right]_0^\pi + \frac{2}{3}\int_0^\pi \sin\theta\, d\theta$$

$$P_T = 30\pi\left(\frac{\delta l I}{\lambda d}\right)^2 . 2\pi d^2 . \frac{4}{3} = 80\left(\frac{\pi\delta l I}{\lambda}\right)^2$$

Or

$$P_T = P_{DM} . 2\pi d^2 . \frac{4}{3} = \frac{8}{3}\pi d^2 P_{DM}$$

The radiation resistance is given by $P_T = I^2 R_{rad}$; $R_{rad} = P_T/I^2$

Then $R_{rad} = 80\left(\frac{\pi\delta l}{\lambda}\right)^2$

The directive gain is given by $D(\theta,\varphi) = \frac{P(\theta,\varphi)}{P_A} = \frac{4\pi d^2 P_D}{P_T}$, where $P(\theta,\varphi) = d^2 P_D$ & $P_A = \frac{P_T}{4\pi}$

$$D(\theta,\varphi) = \frac{4\pi d^2 P_{DM}\, sin^2\theta}{\frac{8}{3}\pi d^2 P_{DM}} = 1.5\, sin^2\theta$$

$D_M = 1.5$

Effective area of the Hertzian dipole = $\frac{\lambda^2 G_T}{4\pi} = 1.5\frac{\lambda^2}{4\pi}$

Dipole Antenna: Dipole Aerial

The simplest type of antenna consists of two metal rods, and is known as a **dipole**. One of the commonest types of antennas is the **monopole** antenna, consisting of a rod situated vertical to a large metal board that serves as a ground plane. The antenna mounted on vehicles is usually a monopole, with the metal roof of the vehicle serving as the ground plane.

The dipole antenna or dipole aerial is one of the most fundamental antenna elements being used for broadcasting, radio reception; two way radio communications, and much more.

The dipole antenna or dipole aerial is one of the most important forms of RF antenna. The dipole can be used on it's own, or it can form part of a more complicated antenna array.

The dipole aerial or antenna is widely used for a variety of types of radio communication, on its own, or incorporated into many other RF antenna designs where it forms the radiating or driven element for the overall antenna.

The most common form of the dipole antenna is the half wave dipole which gains its name because its length corresponds to an electrical half wavelength. The various different types or variants of the dipole antenna tend to be used in different applications - accordingly it can be seen that the dipole is a very flexible and useful antenna.

Half wave dipole antenna:

The half wave dipole antenna is the one that is most widely used. This type of dipole antenna is resonant, operating at a point where it is an electrical half wavelength long.

The electrical half wavelength in the antenna is slightly shorter than that of a half wavelength in free space because of the effect of the wire in which the wave is travelling.

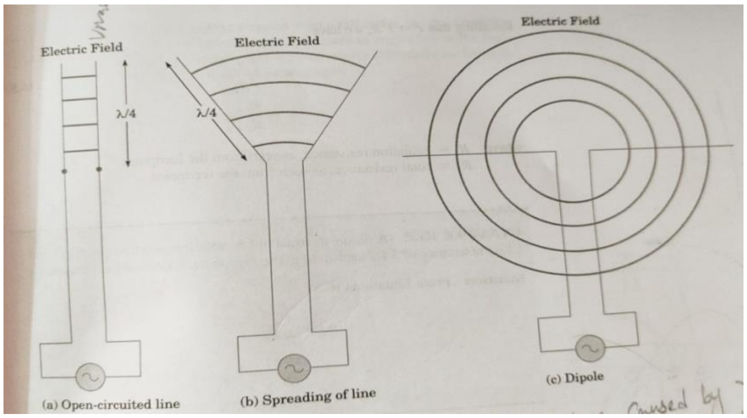
This is the most widely used antenna because of its advantages. It is also known as **Hertz antenna**. A half wave dipole antenna has length equal to the half the wavelength of radiation, but not exactly half, slightly less than one-half. Normally, the length of the half-wave dipole is equal to 95% of one-half the wavelength. The free space wavelength is given by

$$\boldsymbol{\lambda = \frac{c}{f}}$$

Then the length of the antenna is $= \frac{1}{2}(95\% \ of \ \lambda) = \frac{1}{2} \times \frac{95}{100} \times \frac{c}{f} = 0.475 \times \frac{3\times 10^8}{f} = \frac{142.5\times 10^6}{f}$

If frequency is in megahertz, then we can calculate the length of the antenna by using the formula $\frac{142.5}{f}$ (in metres) or $\frac{468}{f}$ (in feets).

If the antenna is open circuited then the electric field flows as shown in fig a. Suppose the two conductors are separated at a point, one quarter wavelength from the end, then the electric field seems to stretch away from the wires (fig b). If the process continues, the electric field detaches itself from antenna and helps to form electromagnetic waves that propagate through the space (fig c).



The half wave dipole is a resonant antenna, the total length of which is nominally λ/2 at the carrier frequency. The spacing between a standing wave maximum and minimum is λ/4 and the current must be zero at the open circuit.

- It will be a maximum λ/4 from the end.
- The voltage is maximum at the end, going to a minimum at the λ/4 point.

A 180° phase shift also occurs along λ/4 section (π/2 for incident wave and π/2 for reflected wave in the opposite direction). At the feed point, the voltage is assumed to be zero. Practically voltage at the feed point will be finite and also the amplitude distribution cannot be same to that of the transmission line section. But these assumptions made the result agree very well with the measured results.

We can consider the half wave dipole as a large number of Hertzian dipole connected in series. There will be a phase difference between radiations from different elements on the half wave dipole as a result of the difference in distance (d_o-d)

By applying Cosine law $OP^2 = PQ^2 + OQ^2 - 2PQ\ OQ\ \cos\theta$

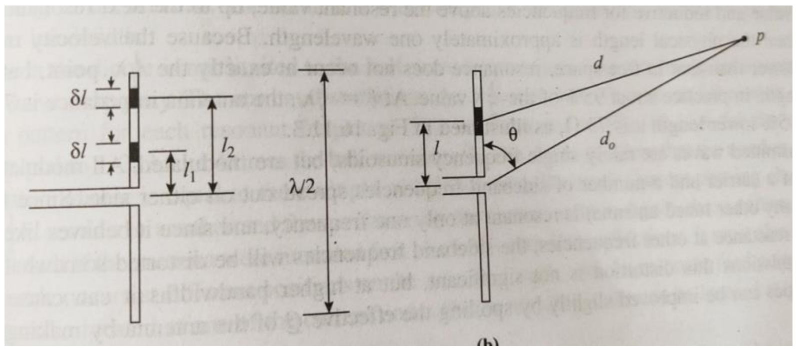
$d^2 = d_O^2 + l^2 - 2d_O l \cos\theta$

$\boldsymbol{d_O \gg l}$, we can neglect l^2

Therefore, $d^2 = d_O^2 - 2d_O l \cos\theta$

$$\boldsymbol{d_O^2 - d^2 = 2d_O l \cos\theta}$$

$$\boldsymbol{(d_O + d).(d_O - d) = 2d_O l \cos\theta}$$



$d_o \approx d$, then $\boldsymbol{(d_O + d) = 2d_O}$

Thus, $\boldsymbol{2d_O(d_O - d) = 2d_O l \cos\theta}$

$$\boldsymbol{(d_O - d) = l \cos\theta}$$

From the expression, it is clear that d_o-d depends on both l and θ. The point 'P' receives response from all Hertzian dipoles and the total response can be found by integrating the individual fields for the full length of the half wave dipole.

Therefore, Peak field strength $\boldsymbol{E_O = \frac{60I_O}{d}F(\theta)}$

Where $\boldsymbol{F(\theta) = \frac{\cos(\pi/2\cos\theta)}{\sin\theta}}$

Thus the normalized gain function becomes, $\boldsymbol{g(\theta) = F^2(\theta)}$

For Hertzian dipole, g(φ) = 1, because of symmetry.

For half wave dipole,

$$l_{eff} = \lambda/\pi,$$

$$D_M = 1.64$$

$$A_{eff} = 0.13\ \lambda^2$$

$$R_{rad} = 73\Omega$$

The total impedance of half wave dipole is a function of frequency. Below resonant frequency, the impedance is completely offered by capacitance while above resonant frequency, impedance will be totally inductive. At l = λ/2, the antenna impedance is Z_A= 73+j42.5

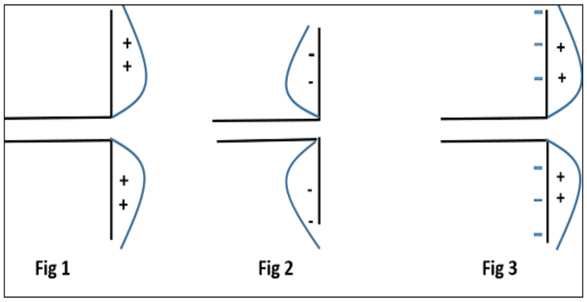
Frequency range

The range of frequency in which half-wave dipole operates is around 3KHz to 300GHz. This is mostly used in radio receivers.

Construction & Working of Half-wave Dipole

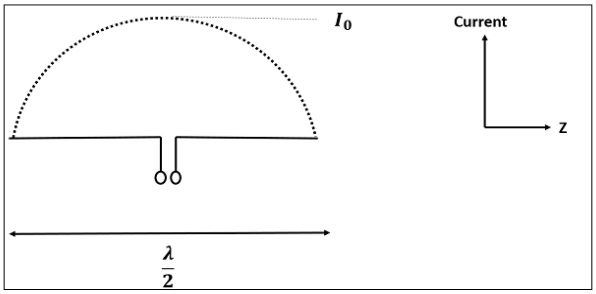
It is a normal dipole antenna, where the frequency of its operation is **half of its wavelength**. Hence, it is called as half-wave dipole antenna.

The edge of the dipole has maximum voltage. This voltage is alternating (AC) in nature. At the positive peak of the voltage, the electrons tend to move in one direction and at the negative peak, the electrons move in the other direction.



The figures given above show the working of a half-wave dipole.

- Fig 1 shows the dipole when the charges induced is in positive half cycle. Now the electrons tend to move towards the charge.
- Fig 2 shows the dipole with negative charges induced. The electrons here tend to move away from the dipole.
- Fig 3 shows the dipole with next positive half cycle. Hence, the electrons again move towards the charge.
- The cumulative effect of this produces a varying field effect which gets radiated in the same pattern produced on it. Hence, the output would be an effective radiation following the cycles of the output voltage pattern. Thus, a half-wave dipole **radiates effectively**.



The above figure shows the current distribution in half wave dipole. The directivity of half wave dipole is 2.15dBi, which is reasonably good. Where, 'i' represents the isotropic radiation.

Advantages

The following are the advantages of half-wave dipole antenna −

- Input impedance is not sensitive.
- Matches well with transmission line impedance.
- Has reasonable length.
- Length of the antenna matches with size and directivity.

Disadvantages

The following are the disadvantages of half-wave dipole antenna −

- Not much effective due to single element.
- It can work better only with a combination.

Applications

The following are the applications of half-wave dipole antenna −

- Used in radio receivers.
- Used in television receivers.
- When employed with others, used for wide variety of applications.

VSWR & Reflected Power

"The ratio of the maximum voltage to the minimum voltage in a standing wave is known as **Voltage Standing Wave Ratio**."

If the impedance of the antenna, the transmission line and the circuitry do not match with each other, then the power will not be radiated effectively. Instead, some of the power is reflected back.

The key features are −

- The term, which indicates the impedance mismatch is **VSWR**.
- **VSWR** stands for Voltage Standing Wave Ratio. It is also called as **SWR**.
- The higher the impedance mismatch, the higher will be the value of **VSWR**.
- The ideal value of VSWR should be 1:1 for effective radiation.
- Reflected power is the power wasted out of the forward power. Both reflected power and VSWR indicate the same thing.

Bandwidth

"A band of frequencies in a wavelength, specified for the particular communication, is known as **bandwidth**."

The signal when transmitted or received is done over a range of frequencies. These particular ranges of frequencies are allotted to a particular signal, so that other signals may not interfere in its transmission.

- **Bandwidth** is the band of frequencies between the higher and lower frequencies over which a signal is transmitted.
- The bandwidth once allotted, cannot be used by others.
- The whole spectrum is divided into bandwidths to allot to different transmitters.

The bandwidth, which we just discussed, can also be called as **Absolute Bandwidth**.

Percentage Bandwidth

"The ratio of absolute bandwidth to the center frequency of that bandwidth can be termed as **percentage bandwidth**."

The particular frequency within a frequency band, at which the signal strength is maximum, is called as **resonant frequency**. It is also called as **center frequency (fC)** of the band.

- The higher and lower frequencies are denoted as $\mathbf{f_H}$ **and** $\mathbf{f_L}$ respectively.
- The absolute bandwidth is given by- $\mathbf{f_H - f_L}$.
- To know how wider the bandwidth is, either **fractional bandwidth** or **percentage bandwidth** has to be calculated.

The **Percentage bandwidth** is calculated to know how much frequency variation either a component or a system can handle.

$$Percentage\ bandwidth = \frac{absolute\ bandwidth}{center\,frequency} = \frac{f_H - f_L}{f_c}$$

Where

- f_H is higher frequency
- f_L is lower frequency
- f_C is center frequency

The higher the percentage bandwidth, the wider will be the bandwidth of the channel.

Radiation Intensity

"**Radiation intensity** is defined as the power per unit solid angle"

Radiation emitted from an antenna which is more intense in a particular direction, indicates the maximum intensity of that antenna. The emission of radiation to a maximum possible extent is nothing but the radiation intensity.

Radiation Intensity is obtained by multiplying the power radiated with the square of the radial distance.

$$U = r^2 \times W_{rad}$$

Where

- **U** is the radiation intensity
- **r** is the radial distance
- $\mathbf{W_{rad}}$ is the power radiated.

The above equation denotes the radiation intensity of an antenna. The function of radial distance is also indicated as **Φ**. The unit of radiation intensity is Watts / steradian or Watts / radian^2.

Radiation intensity of an antenna is closely related to the direction of the beam focused and the efficiency of the beam towards that direction.

Aperture Efficiency

"**Aperture efficiency** of an antenna is the ratio of the effective radiating area (or effective area) to the physical area of the aperture."

An antenna has an aperture through which the power is radiated. This radiation should be effective with minimum losses. The physical area of the aperture should also be taken into consideration, as the effectiveness of the radiation depends upon the area of the aperture, physically on the antenna.

$$\varepsilon_A = \frac{A_{eff}}{A_p}$$

where

- ε_A is Aperture Efficiency.
- A_{eff} is effective area.
- A_p is physical area.

Antenna Efficiency

"**Antenna Efficiency** is the ratio of the radiated power of the antenna to the input power accepted by the antenna." Simply, an Antenna is meant to radiate power given at its input, with minimum losses. The efficiency of an antenna explains how much an antenna is able to deliver its output effectively with minimum losses in the transmission line. This is otherwise called as **Radiation Efficiency Factor** of the antenna.

$$\eta_e = \frac{P_{rad}}{P_{input}}$$

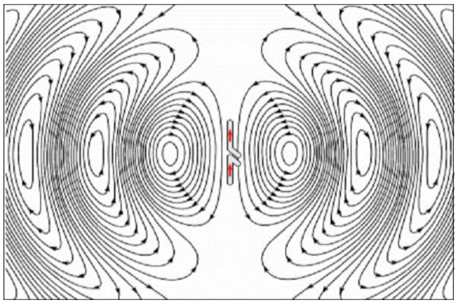
Where

- η_e is the antenna efficiency.
- P_{rad} is the power radiated.
- P_{input} is the input power for the antenna.

Radiation Pattern

The energy radiated by an antenna is represented by the **Radiation pattern** of the antenna. Radiation Patterns are diagrammatical representations of the distribution of radiated energy into space, as a function of direction.

The figure given above shows the radiation pattern of a dipole antenna. The energy being radiated is represented by the patterns drawn in a particular direction. The arrows represent directions of radiation.

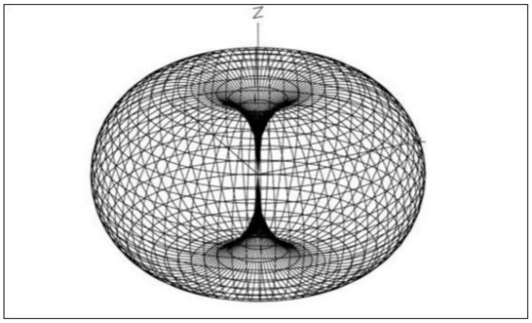


The radiation patterns can be field patterns or power patterns.

- The **field patterns** are plotted as a function of electric and magnetic fields. They are plotted on logarithmic scale.
- The **power patterns** are plotted as a function of square of the magnitude of electric and magnetic fields. They are plotted on logarithmic or commonly on dB scale.

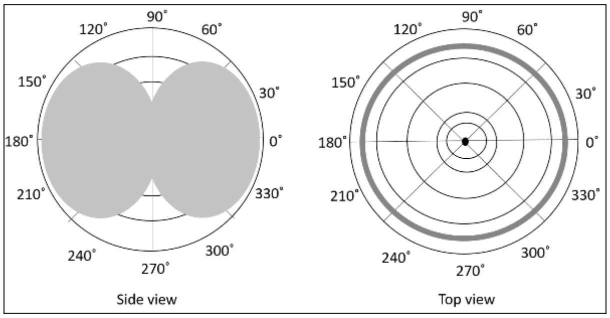
Radiation Pattern in 3D

The radiation pattern is a three-dimensional figure and represented in spherical coordinates (r, θ, Φ) assuming its origin at the center of spherical coordinate system. It looks like the following figure



The given figure is a three dimensional radiation pattern for an **Omni directional pattern**. This clearly indicates the three co-ordinates (x, y, z).

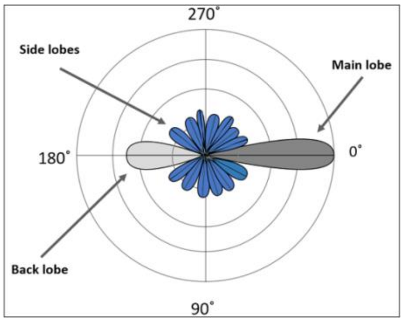
Radiation Pattern in 2D



Two-dimensional pattern can be obtained from three-dimensional pattern by dividing it into horizontal and vertical planes. These resultant patterns are known as Horizontal pattern and Vertical pattern respectively.

The figures show the Omni directional radiation pattern in H and V planes as explained above. H-plane represents the Horizontal pattern, whereas V-plane represents the Vertical pattern.

Lobe Formation



In the representation of radiation pattern, we often come across different shapes, which indicate the major and minor radiation areas, by which the **radiation efficiency** of the antenna is known.

To have a better understanding, consider the following figure, which represents the radiation pattern of a dipole antenna.

Here, the radiation pattern has main lobe, side lobes and back lobe.

- The major part of the radiated field, which covers a larger area, is the **main lobe** or **major lobe**. This is the portion where maximum radiated energy exists. The direction of this lobe indicates the directivity of the antenna.
- The other parts of the pattern where the radiation is distributed side wards are known as **side lobes** or **minor lobes**. These are the areas where the power is wasted.
- There is other lobe, which is exactly opposite to the direction of main lobe. It is known as **back lobe**, which is also a minor lobe. A considerable amount of energy is wasted even here.

Example

If the antennas used in radar systems produce side lobes, target tracing becomes very difficult. This is because, false targets are indicated by these side lobes. It is messy to trace out the real ones and to identify the fake ones. Hence, **elimination** of these **side lobes** is must, in order to improve the performance and save the energy.

Remedy

The radiated energy, which is being wasted in such forms needs to be utilized. If these minor lobes are eliminated and this energy is diverted into one direction (that is towards the major lobe), then the **directivity** of the antenna gets increased which leads to antenna's better performance.

Types of Radiation patterns

The common types of Radiation patterns are −

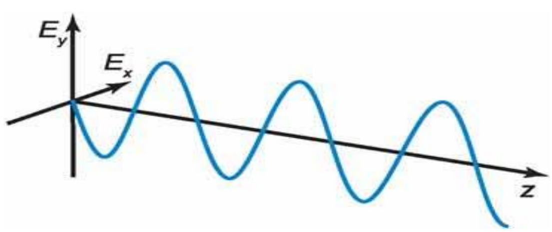
- Omni-directional pattern (also called non-directional pattern): The pattern usually has a doughnut shape in three-dimensional view. However, in two-dimensional view, it forms a figure-of-eight pattern.
- Pencil-beam pattern − The beam has a sharp directional pencil shaped pattern.
- Fan-beam pattern − The beam has a fan-shaped pattern.
- Shaped beam pattern − The beam, which is non-uniform and patternless is known as shaped beam.

A referential point for all these types of radiation is the isotropic radiation. It is important to consider the isotropic radiation even though it is impractical.

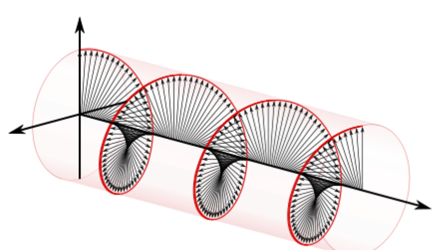
Antenna Polarization

An Antenna can be polarized depending upon our requirement. In the far –field zone, the polarization of the wave is defined by the direction of the electric field vector in relation to the direction of propagation. It can be linearly polarized or circularly polarized. The type of antenna polarization decides the pattern of the beam and polarization at the reception or transmission.

Linear polarization



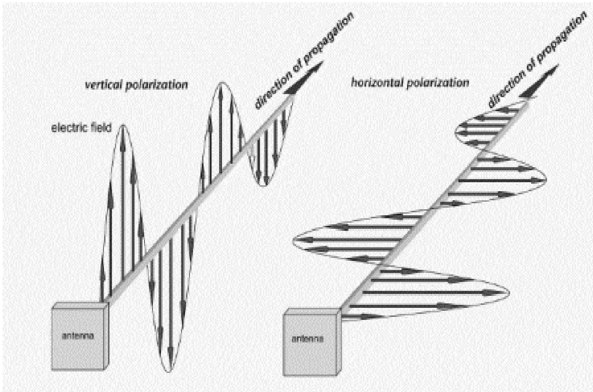
When a wave is transmitted or received, it may be done in different directions. The **linear polarization** of the antenna helps in maintaining the wave in a particular direction, avoiding all the other directions. Though this linear polarization is used, the electric field vector stays in the same plane. Hence, we use this linear polarization to improve the **directivity** of the antenna.



Circular polarization

When a wave is circularly polarized, the electric field vector appears to be rotated with all its components loosing orientation. The mode of rotation may also be different at times. However, by using **circular polarization**, the effect of multi-path gets reduced and hence it is used in satellite communications such as **GPS**.

Horizontal polarization



Horizontal polarization makes the wave weak, as the reflections from the earth surface affect it. They are usually weak at low frequencies below 1GHz. **Horizontal polarization** is used in the transmission of **TV signals** to achieve a better signal to noise ratio.

Vertical polarization

The low frequency vertically polarized waves are advantageous for ground wave transmission. These are not affected by the surface reflections like the horizontally polarized ones. Hence, the **vertical polarization** is used for **mobile communications**.

Each type of polarization has its own advantages and disadvantages. A RF system designer is free to select the type of polarization, according to the system requirements.

Effective Length

Antenna Effective length is used to determine the polarization efficiency of the antenna.

Definition− "The **Effective length** is the ratio of the magnitude of voltage at the open terminals of the receiving antenna to the magnitude of the field strength of the incident wave front, in the same direction of antenna polarization."

When an incident wave arrives at the antenna's input terminals, this wave has some field strength, whose magnitude depends upon the antenna's polarization. This polarization should match with the magnitude of the voltage at receiver terminals.

$$l_e = \frac{V_{oc}}{E_i}$$

Where

- l_e is the effective length.
- V_{oc} is open-circuit voltage.
- E_i is the field strength of the incident wave.

Summary

- ✶ The definition of antenna which converts electrical power into electromagnetic waves and vice versa and the use of antenna either as a transmitting antenna or a receiving antenna is given.
- ✶ The main characteristics of antenna such as its directionality, efficiency and bandwidth are explained. In a radio communication link, the transmitting antenna is coupled to the receiving antenna through the electromagnetic wave.
- ✶ The equivalent circuit of a simple antenna is given and expression for power transmitted and power received is derived.
- ✶ The fictitious resistance termed as radiation resistance is described. The gain and directivity of the antenna is calculated.

- ✶ The concept of effective area and effective length is also explained.
- ✶ The most widely used antenna type half wave dipole which gains its name because its length corresponds to an electrical half wavelength is explained. The construction, working, advantages, disadvantages and applications of half wave dipole are described.
- ✶ The concept of radiation pattern and lobe formation is defined well. Omni-directional pattern (also called non-directional pattern) usually has a doughnut shape in three-dimensional view. However, in two-dimensional view, it forms a figure-of-eight pattern.
- ✶ The different polarization types say linear polarization, circular polarization, horizontal polarization and vertical polarization is discussed.

Review Questions

Two Marks

1. Justify – Antenna is a transducer.
2. What are the characteristics of an antenna
3. Define bandwidth.
4. Define efficiency.
5. Define radiation resistance.
6. Define radiation intensity.
7. What is defined as antenna efficiency?
8. What is known as aperture efficiency?
9. Define radiation power.
10. What are the uses of antenna?
11. What is gain?
12. What are the types of polarization?
13. Define directivity.
14. Define impedance.
15. What is effective length?

Five Marks

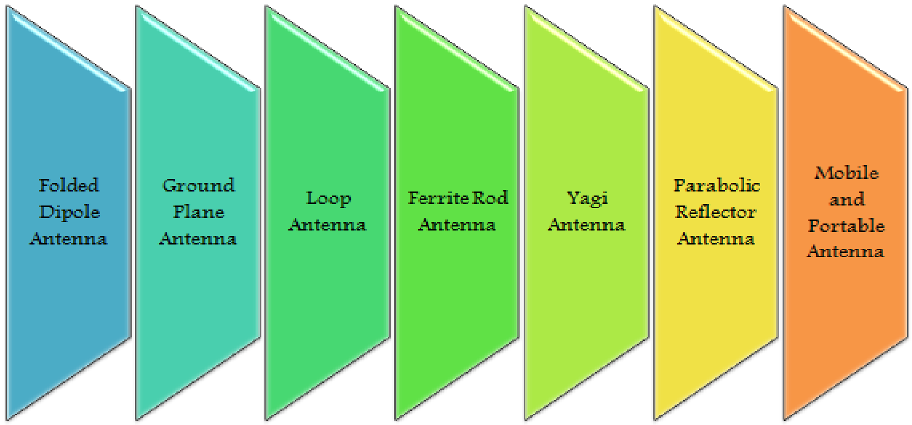
1. Explain the concept of radiation resistance.
2. Explain gain and directivity of an isotropic radiator.
3. Discuss the concept of effective area and effective length.
4. Explain radiation pattern.
5. Discuss antenna polarization.

Ten Marks

1. With necessary diagrams, explain the equivalent circuit of transmitting and receiving antenna.
2. Explain in detail the construction and working of half wave dipole.

Unit III
Types of Antenna

Folded Dipole Antenna – Ground Plane Antenna – Loop Antenna – Ferrite rod receiving antenna – Yagi Array antenna – VHF-UHF Antenna - Parabolic Reflector Antenna – Cell Cite Antenna – Mobile and Portable Antenna.



Introduction

An antenna is used to radiate electromagnetic energy efficiently and in desired directions. Antennas act as matching systems between sources of electromagnetic energy and space. The goal in using antennas is to optimize this matching. The half wave dipole is simple, useful and very common. But there are some other simple antennas in use. Antennas have to be classified to understand their physical structure and functionality more clearly. There are many types of antennas depending upon the applications.

Folded Dipole

Folded dipole has a same length as a standard half wave dipole, but it is made with two parallel conductors, joined at both ends and separated by a distance that is short compared with the length of the antenna. Folded dipole has a wider bandwidth. For this reason, it is used alone or with other elements for TV and FM broadcast receiving antennas. It has high feed point impedance (four times greater than the ordinary dipole).

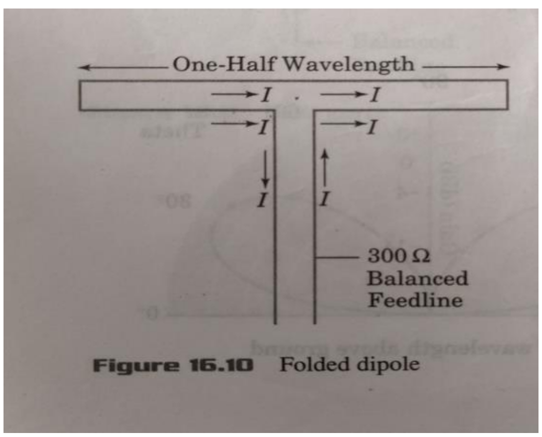


Figure 16.10 Folded dipole

Let V and I be the voltage and current applied to an ordinary dipole. At resonance, feed point impedance is resistive. The power supplied is

$$P = VI$$

where P – average power, V – RMS voltage, I – RMS current.

For folded dipole, the currents in the two conductors will be equal in magnitude as the length in one – half wavelength, magnitude is same but out of phase. As the wire has been folded, the two currents actually flow in the same direction in space and contribute equally to the radiation from the antenna.

If folded dipole and ordinary dipole radiate the same power, the total current must also be the same. But, the current at the feed point of the folded dipole is only one-half the total current. If the feed point current is reduced by one-half, yet the power should remain be the same, then the feed point voltage must be doubled.

i.e., as I = I/2, then $P = \frac{I}{2} 2V$ as $P = VI$

By Ohm's law, resistance at the feed point of the ordinary dipole is $R = \frac{V}{I}$

On substituting V = 2V and I = I/2

$$R' = \frac{2V}{I/2} = 4\frac{V}{I} = 4R$$

Thus the feed point resistance is four times greater than the ordinary dipole.

Ground Plane Antennas

At VHF and above, antenna height is very important, as line of sight propagation is the most common method. An antenna which is mounted at ground level would be ineffective.

It is possible to preserve the simplicity and low radiation angle of the ground mounted monopole by constructing an artificial ground at the base of the antenna. This ground plane can be a conductive sheet, constructed of four or more metal rods radiating outward from the base of the antenna and made as the antenna itself. Ground plane antennas are often seen with CB (Citizen's Band) base stations. As the wavelength in the 27MHz CB band is about 11m, the quarter wave antenna (monopole) needs to be almost 3m long. Such an antenna is obviously difficult enough to mount on a tower, a full-length dipole at twice the length would be much clumsier.

Mobile antennas are usually ground-plane antennas, with the car itself acting as the ground plane. Thus the whip antenna on an automobile would be expected to be one-quarter wavelength in length. Ground plane antenna is

- Practical in FM broadcast band (as the wavelength is about 3m).
- Impractical in CB band (though usable in 11m).
- Out of the question in AM broadcast band (wavelength = 300m).

Loop Antennas:

This antenna is bidirectional, with its greatest sensitivity in the plane of the loop. This would be a multiturn coil for the AM broadcast band.

Ferrite "loopstick" antenna is found in every AM broadcast receiver except in automobiles. Directionality is in the plane of the individual coil turns. The loop antenna is made up of one or more turns of wire on frame (circular or rectangular) and is very much smaller than one wavelength across. This antenna is popular for two reasons.

1) It is relatively compact, lending itself to use with portable receivers.
2) It is quite directive, lending itself to use with direction – finding equipment.

The physical shape of the loop does not affect the radiation pattern except when the length and width of the loop are unequal, in which case a squashed doughnut shape occurs.

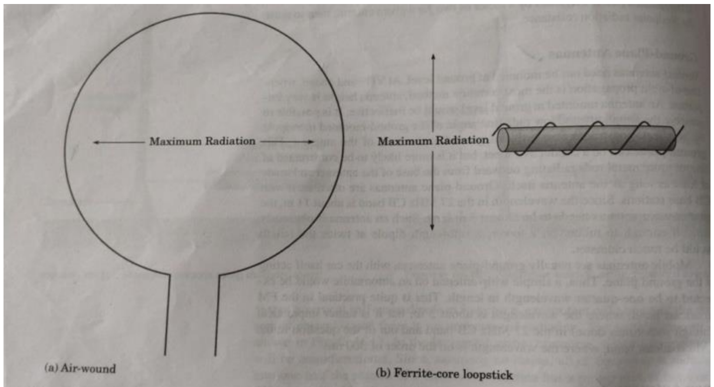


Figure 16.14 Loop antennas

Loop antennas made of several turns of wire around a rectangular frame were popular for earlier model broadcast receivers. Recently, these have been almost entirely replaced by the smaller ferrite – rod antennas. When the loop is aligned for maximum signal strength, the maximum flux linkages are **BAN**

where B- RMS magnetic flux, A – loop area and N – number or turns in the loop.

By Faraday's law, the induced emf is given by

$$\mathbf{V_s = \omega BAN}$$

ω – angular frequency

When the loop is turned by means of an external capacitor to the received frequency, the voltage at te capacitor terminal is magnified by the circuit Q. Then

$$\mathbf{V_{max} = V_s Q = \omega BANQ}$$

The Q is determined by the desired selectivity

- ♣ The area must be kept small, increasing the number of turns increases the coil inductance and changes the Q and even changing the flux density 'B' affects the Q.
- ♣ However changing the flux density by using a magnetic core can be achieved by minimal change of Q using ferrite cores.

Ferrite – Rod Antenna

The ferrite rod antenna is made by winding a coil of wire on a ferrite rod. Ferrites are the material that exhibits the properties of ferromagnetism. These materials exhibit a high relative permeability same as that of the magnetic metals, but unlike ferromagnetic materials, they also have a high bulk resistivity.

This means that at high frequencies, eddy currents induced within the materials are practically nonexistent and high Q-coils can be made. Typical values of N is around 100 and resistivity around 10,000 Ωcm. A high length to diameter ratio for the rod gives a desired high permeability.

The size of the coil plays the important role. If the coil is too long compared to the rod length, the change of permeability with the temperature will cause a noticeable change in the inductance. If it is too short, the Q will be low. Positioning the coil on the core is difficult (complicate), since the effective permeability is a function of position on the rod, ranging from a maximum at the centre to a minimum at either end. When more than one coil is mounted on the same rod, they must be placed at opposite ends to minimize interactions between them.

The coil of wire on the ferrite rod is basically a modified loop antenna, so the maximum induced emf appearing at its terminals is given by

$$\mathbf{V_s = \omega BANF\mu_r}$$

F – Modifying factor accounting for coil length, ranging from unity for short coils to about 0.7 for one that extends the full length of the rod.

μ_r – Effective relative permeability of the rod.

A – Cross sectional area of the rod

An expression for effective length of a ferrite rod can be given as

$$\mathbf{V_A = El_{eff} \Rightarrow l_{eff} = \frac{V_A}{E}}$$

$$\mathbf{l_{eff} = \frac{V_A}{v_p B} = \frac{V_A}{\lambda f B}}$$

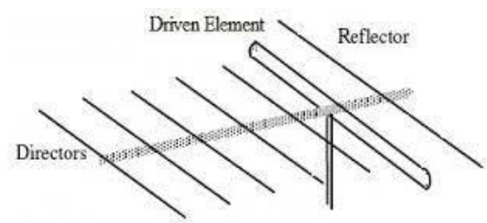
$$\mathbf{l_{eff} = \frac{2\pi f BANF\mu_r}{\lambda f B}}$$

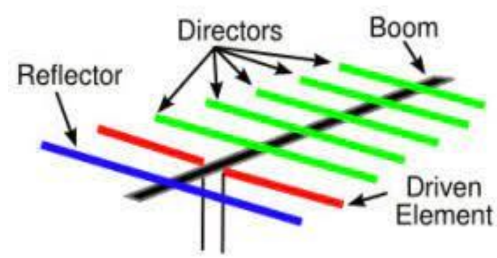
$$\mathbf{l_{eff} = \frac{2\pi ANF\mu_r}{\lambda}}$$

Since the voltage appearing at the terminals is of more importance in receiving antenna, the factor Ql_{eff} is often given as a figure of merit for rod antennas. The directional properties of the ferrite rod antenna are similar to those of the loop antenna.

Yagi Array

The Yagi array has one driven element, one reflector behind the driven element and one or more directors in front of the driven element. The driven element is a half wave dipole or folded dipole. The reflector is slightly longer than the driven element and directors are shorter than the driven element. The spacing between elements varies. The Yagi antenna is more formally referred to as the **Yagi Uda array**.

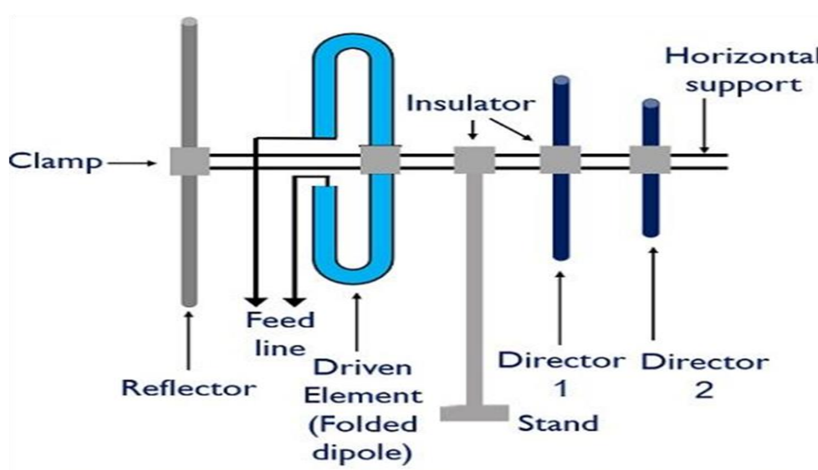




Yagi-Uda antenna was named after two scientists from Japan, who developed and explained this antenna. Professor **Shintaro Uda** initially explained the theory of this antenna in the Japanese language in the year 1928. However, later this antenna was explained in English by **Hidetsugu Yagi**. It is sometimes known as Aerial or Yagi Antenna. Though this was developed in the year 1928 but is adopted for personal and commercial use since **1930**.

Construction

A Yagi-Uda antenna has 3 main elements that jointly form its structure. These 3 major elements are driven element which is generally a half-wave folded dipole, a reflector and directors. The figure below represents the structure of the Yagi-Uda antenna:



Basically, the arrangement is said to be an array of **active** and **parasitic elements**. The dipole generally a metallic rod acts as the active element as external feeding is provided to it using transmission lines, while reflector and directors are the parasitic elements of the structure. The parasitic elements are also metallic rods placed parallel in the line of sight orientation with respect to the driven element. It is noteworthy here that no external excitation is provided to the parasitic elements. All these elements are mounted on a centre rod that acts as horizontal support.

The reflector is present at one of the ends of the metallic rod and has length around, **5% greater** than the length of the driven element. While the directors are almost **5% shorter** than the driven element (i.e., λ/2 at the resonant frequency) and are placed at the other side of the dipole as these are used to provide maximum directivity to the antenna.

So, for 3 element aerial, the lengths of the elements can be considered as:

$$Length\ of\ driven\ element = \frac{475}{f_{MHz}}\ feet$$

$$Length\ of\ reflector = \frac{500}{f_{MHz}}\ feet$$

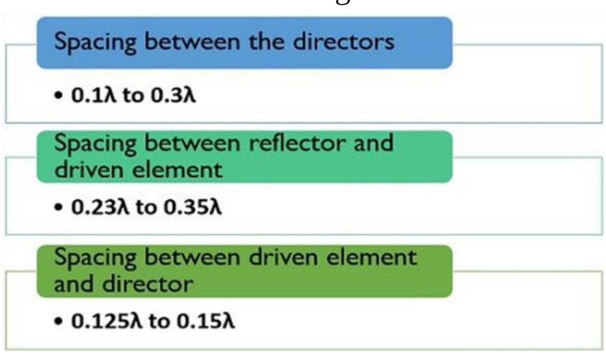
$$Length\ of\ director = \frac{455}{f_{MHz}}\ feet$$

Working of Yagi-Uda Antenna

The element having a length greater than λ/2 i.e., the reflector, shows inductive characteristic, therefore, the current in the reflector lags the induced voltage whereas the director is capacitive. So, the current flowing through it leads the voltage.

As the director is placed in front of the driven elements, directors add the field of the driven element in the direction away from it. When multiple directors are placed in the arrangement then each director will provide excitation to the next one. Also, the reflector in the opposite direction as that of the director when accurately placed adds the field in the direction towards the driven element. This is done in order to reduce the losses due to the back radiated wave as much as possible. In order to get the **additional gain**, multiple directors can be used in the direction of the beam.

The spacing between the elements to form a Yagi -Uda structure is as follows:



Basically, the induced voltage and the current flowing due to the induced voltage in the element vary with the spacing between the active and parasitic elements along with the reactance associated with the elements.

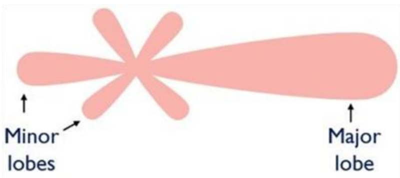
It is to be noted here that with the increase in distance between driven element and director, there will be more need for capacitive reactance in order to provide accurate phasing to the current in the director. Thus, the length of the director is kept small to get the capacity reactance.

Characteristics

- A Yagi-Uda antenna is said to be **beam antenna** if it is only 3 elements array i.e., a driven and reflector and only a single director.

- It offers moderate unidirectional directivity.
- The gain offered by the Yagi-Uda antenna is around **8 dB** with front to back ratio of approximately **20 dB**.
- In order to increase the directivity, more elements can be added in the array.
- Another name to this antenna is **super directive** antenna due to its high directive gain.
- It is **frequency sensitive** thus is a fixed frequency device.

Radiation Pattern of Yagi-Uda Antenna



Here the major lobe represents the forward radiated wave while the major lobe represents the back radiated wave.

Advantages

1. Yagi-Uda antenna offers very high gain.
2. It possesses a highly directional characteristic because of the use of directors.
3. It is a low-cost antenna.
4. Yagi-Uda antenna shows suitability towards high-frequency operations.
5. It is light in weight and feeding mechanism is also simple.
6. It is power efficient.
7. It also offers ease of construction and handling.

Disadvantages

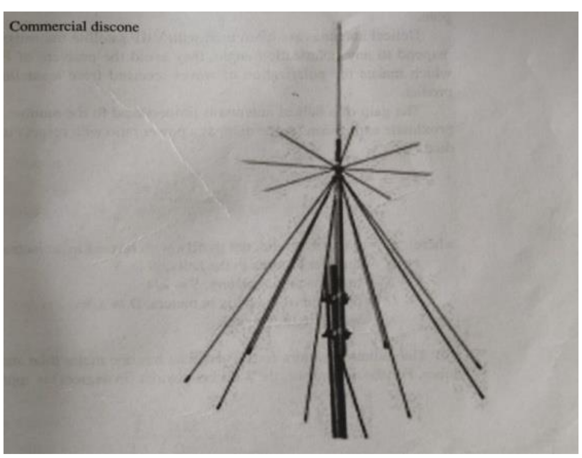
1. These antennas are highly affected by atmospheric conditions.
2. Noise is the major factor that disturbs the overall performance of the antenna.

Applications of Yagi-Uda Antenna

- In TV signal reception, as it has excellent receiving ability.
- Astronomical and defense related applications
- In radio astronomy

VHF – UHF Antennas

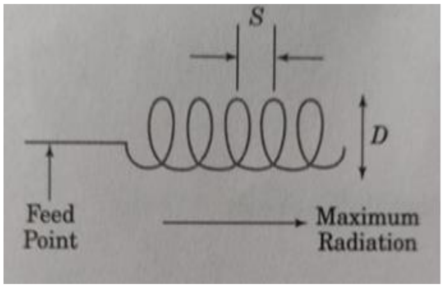
Discone Omni



The discone antenna is designed to radiate an omnidirectional pattern in the horizontal plane, with vertical polarization. It is characterized by very wide bandwidth, covering approximately a 10:1 frequency range. The feed point impedance is approximately 50Ω. The feed point is located at the intersection of the disk and the cone. The disk-cone combination acts as a transformer to match the feedline impedance to the free space impedance, which is 377Ω.

The wide bandwidth of the discone makes it a very popular antenna for general reception in the VHF and UHF ranges. It is a favourite for use with scanners. This type of antenna is ideal for base station operation for urban mobile communication systems, since it gives a good omnidirectional pattern, is physically very compact and rugged, and is quite inexpensive to construct. Its directional gain along the horizontal plane is comparable that of the dipole antenna.

Helical Antenna



The radiator element of a helical antenna is basically a coil of wire. A quarter wave monopole antennas can be shortened and wound into a helix (spiral). This is the **common rubber ducky** antenna used with many handheld receivers. Sometimes it is called a helical antenna and it certainly helical in shape.

The helical antenna is used with a plane reflector to improve its directional characteristics. The circumference of each turn is about one wavelength, and the turn are about one-quarter wavelength, and the twins are about one-quarter wavelength. Helical antenna is used to receive circularly polarized waves and also plane polarized waves with the polarization in any direction. The gain is proportional to the number of turns and can be several decibels greater than the dipole.

Helical antennas are often used with VHF satellite transmissions. The gain of a helical antenna is approximately given as a power ratio with respect to an isotropic radiator

$$\mathbf{G = \frac{15NS(\pi D)^2}{\lambda^3}}$$

G – Gain (not in dB) w.r.to an isotropic radiator

N – Number of turns in the helix

S – Turn spacing

D – Diameter of the helix

λ – Wavelength

The radiation pattern for the antenna has one major lobe and several minor lobes. For the major lobe, the 3dB beam width is approximately

$$\boldsymbol{\theta = \frac{52\lambda}{\pi D}\sqrt{\frac{\lambda}{NS}}}$$

There is yet another type of antenna which resembles a half wave dipole except that the ends are bent. This antenna is used to produce both horizontally and vertically polarized waves simultaneously with one half the input powers going into each polarization. Such antennas are very common in FM broadcasting and enable the signal to be received with both vertical antennas (such as those on cars) and horizontal antennas (like the TV antennas installed on the roof of houses). The antenna is approximately omnidirectional in the horizontal plane.

Parabolic Reflector

The most widely used for microwaves is the parabolic reflector, which consists of a primary antenna such as dipole or horn situated at the focal point of a parabolic reflector. Parabolic reflectors have the useful property that any ray originates at a point called a focus and strikes the reflected surface will be reflected parallel to the axis of the parabola (a collimated beam of radiation will be produced).

Ideally the antenna at the feed point should illuminate the entire surface of the dish with the same intensity of radiation and should not spill any radiation off the edges of the dish or in other directions, such that the gain and beam width of the antenna could be easily calculated. The equation for gain is

$$\mathbf{G = \frac{\pi^2 D^2}{\lambda^2}}$$

G – Gain as a power ratio
D – Diameter of the dish
λ – Free space wavelength

The gain may be reduced by uneven illumination of the antenna, losses and any radiation spilling off at the edges. To include these effects in gain calculations, it is necessary to include a constant η, which is known as the efficiency of the antenna. This constant can theoretically have a value between zero and one but is between 0.5 and 0.7 for a typical antenna. Then the equation becomes

$$\mathbf{G = \frac{\eta\pi^2 D^2}{\lambda^2}}$$

The equation for the beam width is

$$\mathbf{\theta = \frac{70\lambda}{D}}$$

θ – Beam width in degrees at the 3dB points
λ – Free space wavelength
D – Diameter of the dish

An isotropic point source is assumed to be situated at the focal point. In addition to the desired parallel beam, it can be seen that some of the ray are not captured by the reflector and these constitute spill over. In the receiver mode, spill over increases noise pick up, which can be particularly troublesome in satellite ground stations. Also some radiation from the primary radiator occurs in the forward direction in addition to the parallel beam. This is termed back lobe radiation since it is from the back lobe of the primary radiator. Back lobe radiation is undesirable because it can interfere destructively with the reflected beam, and practical radiators are designed to eliminate or minimize this.

The isotropic radiator at the focal point will radiate spherical waves and the parabolic reflector converts these to plane waves. The directivity of the parabolic reflector is a function of the primary antenna directivity and the ratio of focal length to reflector diameter, f/D. This ratio is known as the aperture number, determines the angular aperture of the reflector 2ψ, which in turn determines how much of the primary radiation is intercepted by the reflector.

Making f/D too large, increases spill over to the extent that aperture efficiency then decreases. Reducing f/D to less than ¼ places the focal point inside the reflector. Placing the primary antenna too close to the reflector reflects in the reflector affecting the primary antenna impedance and radiation pattern, which is difficult to take into account. It can be shown that the aperture efficiency peaks at about 80%, with the angular aperture ranging from about 40º to 70º depending on the primary radiation pattern.

The relationship between aperture number and angular aperture is

$$\mathbf{\frac{f}{D} = 0.25 \cot\left(\frac{\Psi}{2}\right)}$$

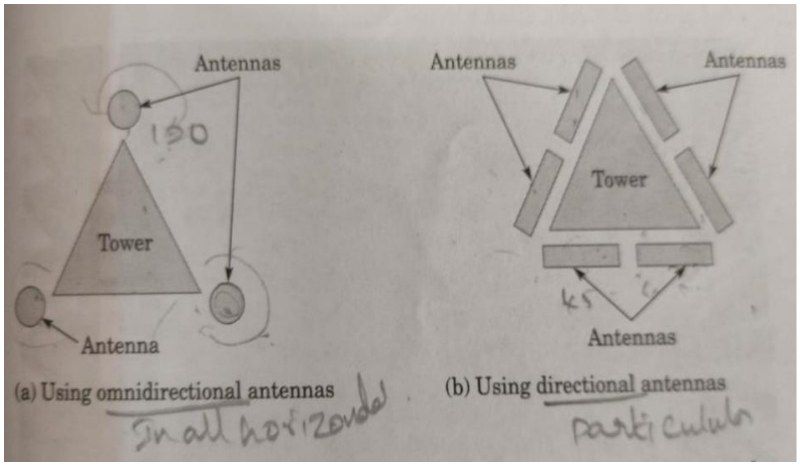
For an angular aperture of 55º, the aperture number is 0.48. As f/D is greater than ¼ , the focal point should lie outside the mouth of the reflector. The effective area is given by

$$\mathbf{A_{eff} = \eta A}$$

where $\mathbf{A} = \frac{\pi D^2}{4}$, A – Physical area of the reflector aperture.

Any type of antenna can be used with a parabolic reflector; a horn antenna provides a simple and efficient method to feed power to the antenna. Besides the simple horn feed, there are several other ways to get power to the parabolic reflector. A feed horn in the centre of the dish itself radiates to a reflector at the focus of the antenna (Gregorian feed). This reflects the signal of the main parabolic reflector. By removing the feed horn from the focus, this system allows any waveguide or electronics associated with the feed point to be placed in a more convenient location.

Cell-Site Antenna



(a) Using omnidirectional antennas (b) Using directional antennas

These are the antennas used for 800 MHz cellular radio and 1900 MHz personal communication service (PSS) system. The antennas used for cellular radio systems have to be omnidirectional and with beam widths of 120º and less for sectorized cells. Narrower beam widths are also useful for filling in dead spots (where wifi does not work). Typical cellular antennas use variation of the collinear antenna for omnidirectional patterns and either collinear antenna backed by reflectors or log periodic antennas for directional patterns.

When cell sizes are small, in high traffic areas, directional base station antennas are often tilted downwards in order to reduce the distance travelled by the signals and the interference to neighbouring cells. The down tilt can be done mechanically and also be done electrically. Cellular and PCS base station receiving antennas are usually mounted in such a way as to obtain space diversity.

For an omnidirectional pattern, typically three antennas are mounted at the corners of a tower with a triangular cross section. When the cell is divided into three 120º sectors, it is usual to mount two antennas for each sector on the sides of the tower. It is possible to use single antenna for both receiving and transmitting using duplexes, but often the transmitting antenna is located separately. Only one transmitting antenna is required for a cell with an omnidirectional pattern, other mounting locations, such as walls of buildings require some variations.

A recent development that can reduce the number of antennas required for diversity is the use of dual polarization antennas. In a cluttered mobile environment, signal polarization may be randomized by reflections. In that case, diversity can be achieved by using two polarizations, at 45ºangles to the vertical. Dual polarization antennas can considerably reduce the number of visible structures needed at a cell site.

Mobile and Portable Antennas

The portable and mobile antennas used with cellular and PCS systems have to be omnidirectional and small, especially in the case of portable phones. It should be easier to achieve at 1900 MHz than at 800 MHz. Many PCS phones must double as 800 MHz cell phones; however these antennas work well at 800 MHz.

The simplest suitable antenna is a quarter wave monopole and these are the usual antennas supplied with portable phones. For mobile phones, where compact size is not quite as important a very common configuration consists of a quarter wave antenna with a half wave antenna mounted collinearly above it. The two are connected by a coil which matches impedances.

Summary

- Folded dipole has a same length as a standard half wave dipole, but it is made with two parallel conductors, joined at both ends and separated by a distance that is short compared with the length of the antenna.
- This ground plane can be a conductive sheet, constructed of four or more metal rods radiating outward from the base of the antenna and made as the antenna itself.
- Loop antenna is bidirectional, with its greatest sensitivity in the plane. This would be a multiturn coil for the AM broadcast band.
- The ferrite rod antenna is made by winding a coil of wire on a ferrite rod. Ferrites are the material that exhibits the properties of ferromagnetism. Ferrite “loopstick” antenna is found in every AM broadcast receiver except in automobiles.
- The Yagi array has one driven element, one reflector behind the driven element and one or more directors in front of the driven element. The driven element is a half wave dipole or folded dipole.
- The most widely used for microwaves is the parabolic reflector, which consists of a primary antenna such as dipole or horn situated at the focal point of a parabolic reflector.
- Parabolic reflectors have the useful property that any ray originates at a point called a focus and strikes the reflected surface will be reflected parallel to the axis of the parabola

Review Questions

Two Marks

1. Write down the advantages of loop antenna
2. Give the benefits of ferrite rod antenna.
3. What is loop antenna?
4. Define VHF antenna.
5. What is meant by ground plane antenna?
6. What is meant by portable antenna?
7. What are the advantages of cell cite antenna
8. What is the use of Yagi Uda antenna?
9. What are the basic components of Yagi antenna?
10. What is parabolic reflector?

Five Marks

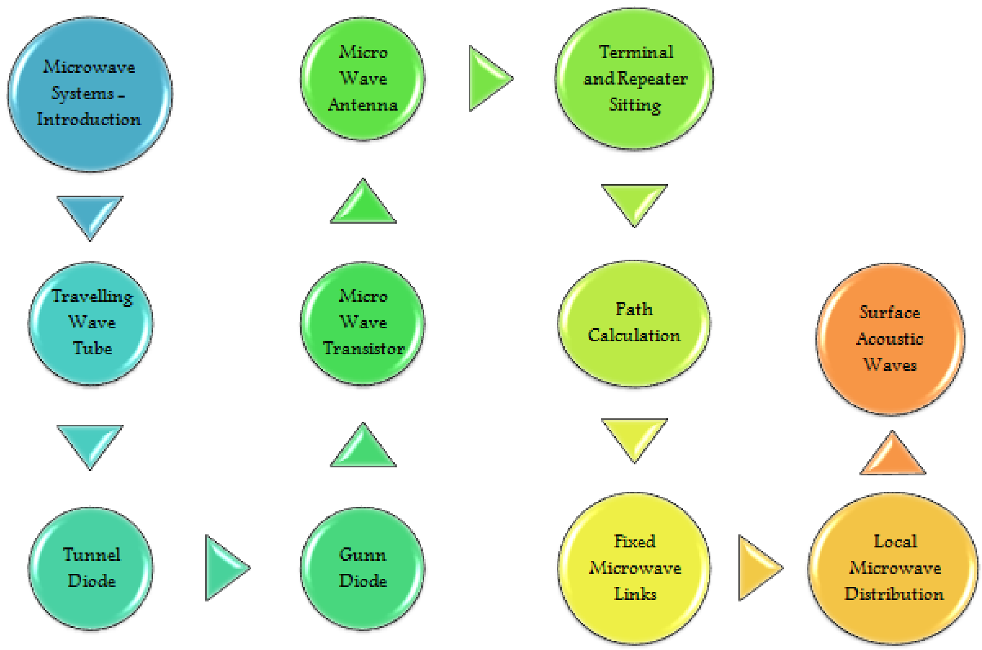
1. Discuss in detail the folded dipole antenna.
2. Explain loop antenna
3. Explain ferrite – rod antenna.
4. Briefly explain helical antenna.
5. Explain cell-cite antenna
6. Discuss about the mobile and portable antenna.

Ten Marks

1. Explain the construction and working of Yagi array antenna.
2. Explain in detail about the parabolic reflector.

Unit IV
Micro Waves System

Introduction to micro waves – Generation of micro waves using Travelling Wave Tube (TWT) – Tunnel Diodes – Gunn Diode – Microwave Transistor – Micro wave antenna (Horn)– Terminal and Repeater Sitting – Path Calculation – Fixed Microwave Links – Local Microwave Distribution – System Surface Acoustic waves.



INTRODUCTION

Microwaves are used in radar, radio transmission and other applications. Microwaves are electromagnetic waves generally defined as lying within the frequency range of 100 MHz – 300 GHz (3m – 1mm in wavelength).

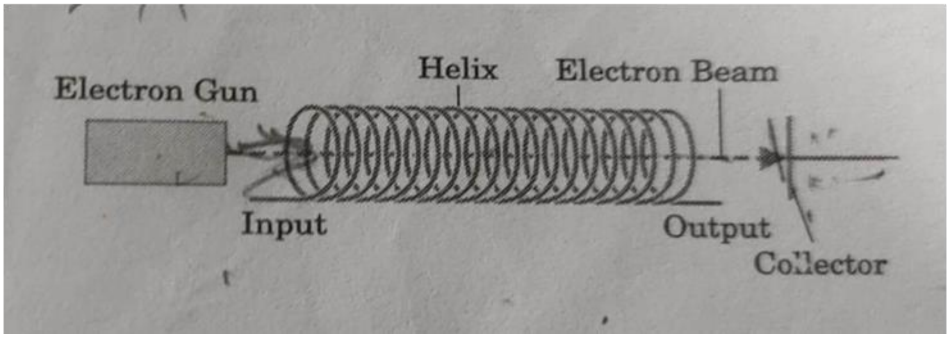
Microwaves are important, because of

- Its excellent and wide use in wireless communication.
- A boon for the transmission and reception of electromagnetic waves by several methods of wave propagation.
- Power efficient transmission is possible
- Excellent heating property
- As antenna size is directly proportional to the wavelength of the electromagnetic waves, using microwaves antenna size is reduced to a great extent and hence proper use of microwave communication make the communication system much simpler than other communication systems.
- Due to reduced antenna size, the design of the portable communication system becomes easier.

TRAVELLING WAVE TUBE (TWT)

TWT can be used as a moderate power amplifier or as an oscillator. It is distinguished by its wide band width. The transmitters in communication satellites and their ground states often use TWT amplifiers. TWT is a linear beam tube. It has a slow wave structure. It may consist of a series of coupled cavities, but it has the form of a helix. The helix can be thought of as a sort of waveguide along which a wave travels by going around and around the turns of the spiral. The progress of the wave along the length of the helix is much slower than its speed as it moves around the turns.

An electron beam travels from the cathode down the centre of the helix to the collector, giving up some of its energy to the wave as it does so. Since the helix unlike a series of cavities is non-resonant, the band width of the TWT can be much greater. On the other hand, it is more difficult to remove heat from the helix. So the helix TWT is a low to medium power device for power levels up to about a kilowatt. The TWT needs a magnetic structure to focus the electron beam so that it passes within the helical slow wave structure.



Traveling Wave Tubes are vacuum tubes used as high-gain, low-noise, wide-bandwidth microwave wideband amplifiers. A TWT is capable of gains from 40 to 70 dB. TWTs have been designed for frequencies as low as 300 Megahertz and as high as 100 Gigahertz. Power level ranges from a few watts to 10 MW. The TWT is primarily a voltage amplifier. Together with the klystrons they form a special group of linear-beam tubes in context of velocity-modulated tubes. There are two different main types of TWT:

- Low-power Helix TWT occurs as a highly sensitive, low-noise and wideband amplifier in radar receivers and measurement equipment.
- High-power Coupled-Cavity TWT are used as a power-amplifier for high-power transmitters, e.g. as pre-amplifier for crossed field amplifiers (CFA). They have significant higher output-power but less bandwidth.

Both types have the same operating principles and they both incorporate the basic components shown in Figure. They mostly differ in the construction of the slow-wave structure. The wide-bandwidth and low-noise characteristics make the TWT ideal for use as an RF amplifier in microwave equipment. On reason of the special low-noise characteristic they are widely used as an active RF amplifier element in microwave receivers and transmitters in radar systems and in space communications.

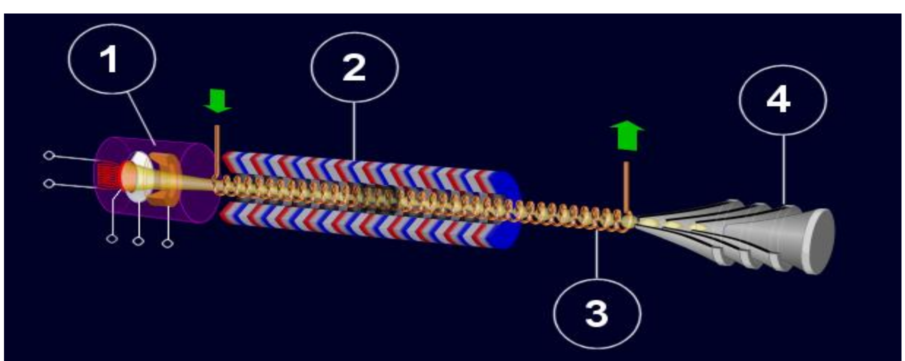


Figure 1: Physical construction of a TWT: ① Electron gun; ② Surrounding magnet; ③ Slow wave structure (here: Helix); ④ Collector; (interactive picture)

Physical construction

1. The physical construction of a typical TWT is shown in Figure 1. It consists of four basic elements.
2. Electron gun which produces and then accelerates an electron beam along the axis of the tube.
3. Magnetic electron beam focusing system which provides a magnetic field along the axis of the tube to focus the electrons into a tight beam.
4. Slow wave structure as RF- interaction circuit, e.g. a coiled wire (Helix) at the center of the tube that provides a low-impedance transmission line for the RF energy within the tube.
5. Collector. The electron beam is received at the collector after it has passed through the slow wave structure.

All components of the TWT are held under a very high vacuum. The RF input and output may couple onto and removed from the helix by waveguide directional couplers that have no physical connection to the helix.

Electron Gun

The electron gun is similar in construction as in all cathode ray tubes. It consists of a indirect heated cathode, that must be heated to a temperature between 850° and 1100° Celsius (1500° to 2000° Fahrenheit) to produce appreciable electron emission. A focusing grid with the same potential as the cathode (or a small negative bias up to −20 Volts relative to the cathode) directs the electrons in the desired direction. One or more anodes are used to generate the requisite electron velocity. The beam passes the anodes through a hole or a grid and travels through the slow wave structure. The electron gun is covered by a shielding box to prevent hazardous radiation.

Surrounding Magnet

The surrounding magnet provides a magnetic field along the axis of the tube to focus the electrons into a tight beam. This magnet may be either a permanent magnet or a solenoid (electromagnetic) focusing element (see Figure 2a). A permanent magnet doesn't need a power supply and ensures that the magnetic field is always present. The disadvantage is that a permanent magnet doesn't provide an adjustment of the magnetic field to optimize the tubes performance.

If a single permanent magnet (see Figure 2b) is replaced by a number of smaller magnets then the size and total weight of the magnet structure is reduced (see Figure 2c).

The housing is usually made of aluminum to prevent the disturbing influence of ferromagnetic materials. Extrinsic magnetic materials can interfere with the uniform magnetic field and destroy the traveling wave tube. Therefore, the packaging of a traveling wave tube has oversized dimensions often.

Slow wave structure

Since the electron beam into the tube must obviously travel slower than the speed of light, there must be some means of slowing down the forward velocity of the electromagnetic wave. The electron beam speed of a TWT is about 10 to 50 percent of the speed of light. The speed depends on the cathode voltage that may be between 4 to 120 Kilovolts. The slowdown is done by means of a slow wave structure, on which the electromagnetic wave propagates.

Collector

The collector is a voltage electrode of the TWT. It's the same potential as the body of the tube, and this is usually on ground. In the absence of an input signal, the entire beam energy must be dissipated in the collector. Forced air-cooling or liquid cooling of the collector is necessary at high-power TWTs. High-power TWTs often use multi-stage collectors as shown in Figure 1.

Functional describing

The input voltage creates an additional axial electric field that moves as fast as the electron beam on the wire of the helix. This electric field accelerates (in the positive half-wave) or decelerates (in the negative half-wave) the electrons in the electron beam. This process is called velocity modulation. If the electrons of the beam were accelerated to travel faster than the waves traveling on the wire, electron bunching would occur through the effect of velocity modulation. (see Figure 4)

By delivering energy to the electron beam, the power of the traveling wave decreases. The additional attenuator causes a decreasing to zero. This one attenuator also prevents any reflected waves from traveling back down the helix.

However, the velocity modulation is still effective in the electron beam. The faster electrons catch up with the slower electrons and bunching occurs. The electron- beam bunching already starts at the beginning of the helix and reaches its highest expression on the end of the helix. The electron bunches in the beam give up energy to the wire of the slow wave structure. They repel the electrons in the wire and generate a new one traveling wave in the helix. The energy from bunches would increase the amplitude of the traveling wave in a progressive action that would take place all along the length of the TWT.

The injection of the wave in the slow wave structure (as shown in Figure 5) causes a phase shift of −90 degrees relative to the initial waveform. When the electrons deliver their energy to the wave in the helix, they slow down. In some TWTs the helix is made narrower at the end of the tube therefore. This slows down the speed of the electromagnetic wave in the slow wave structure as well.

Characteristics of a TWT

Power amplification

The attainable power-amplification is essentially dependent on the following factors:

- constructive details (e.g. structure and length of the helix)
- electron beam diameter (adjustable by the density of the focusing magnetic field)
- power input (see Figure 6)
- voltage U_{A2} on the helix

As shown in Figure 6, the gain of a given TWT has got linear characteristic of about 26 dB at small input power. If you increase the input power, the output power doesn't increase for the same gain. So you can prevent a saturation of e.g the following mixer stage in radar receiver. The relatively low efficiency of the TWT partially offsets the advantages of high gain and wide bandwidth.

Bandwidth

The gain of a TWT is affected by the interaction of the electrons with the electric field caused by the wave in the slow wave structure. The effectiveness depends on the frequency response of the slow wave structure. A helix may have a bandwidth of more than two octaves. If the slow wave structure contains resonant parts, then the bandwidth depends on its frequency response. The bandwidth of commonly used Coupled-Cavity TWTs is about 10 … 20 percent of the center frequency.

Noise Figure

The most important parameter for the use of the traveling wave tube as a pre-amplifier in radar receivers is the noise figure of the traveling wave tube. This determines the sensitivity of the receiver and thus the maximum range of the radar. The noise figure of recently used TWTs is 3 … 10 dB. There are three unavoidable sources of noise in a traveling wave tube:

- Shot noise results from the random emission of electrons of the cathode
- Velocity noise arises from different velocities of the emitted electrons.
- Johnson–Nyquist noise is the electronic noise generated by the thermal agitation of the electrons.

The noise figure depends on the size of most supply voltages of the traveling wave tube. For example, if the voltages at the electrodes are 5% less than the optimum values, the noise figure approximately doubles.

Different Slow Wave Structures

The previously described helix may be replaced by some other slow wave structure such as a ring-bar, ring loop, or coupled cavity structure. The structure is chosen to give the characteristic appropriate to the desired gain/bandwidth and power characteristics.

Contra-wound Helix

A contra-wound helix uses two helices wound in opposite directions. Both helices must be identical in dimensions. A contra-wound helix is less sensitive to backward waves interactions and therefore allows higher operating voltages, currents and power. The penalty for these advantages is that the bandwidth is less than that of a single helix.

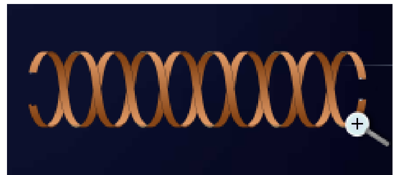


Figure 7: Contra-wound Helix slow wave structure

Ring-Loop TWT

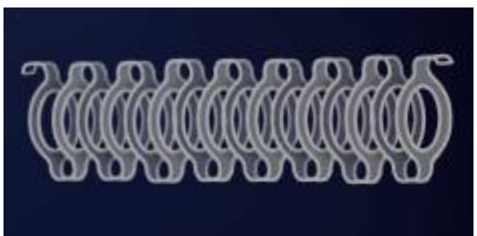


Figure 8: Ring-Loop slow wave structure

A Ring Loop TWT uses loops as slow wave structure to tie the rings together. These devices are capable of higher power levels than conventional helix TWTs, but have significantly less bandwidth of 5…15 percent and lower cut-off frequency of 18 GHz.

The feature of the ring-loop slow wave structure is high coupling impedance and low harmonic wave components. Therefore ring-loop traveling wave tube has advantages of high gain (40…60 Decibels), small dimension, higher operating voltage and less danger of the backward wave oscillation.

Ring-Bar TWT

The Ring-Bar TWT was developed from the contra-wound helix and has got the same characteristics likely the Ring-Loop TWT. This one slow wave structure is very easy to make by precise laser cuts in a thin copper pipe.

Figure 9: Ring-Bar slow wave structure

Coupled-cavity TWT

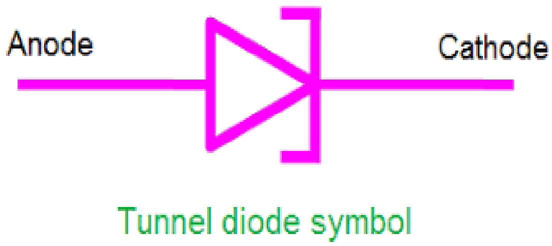
Figure 10: Coupled-cavity slow wave structure

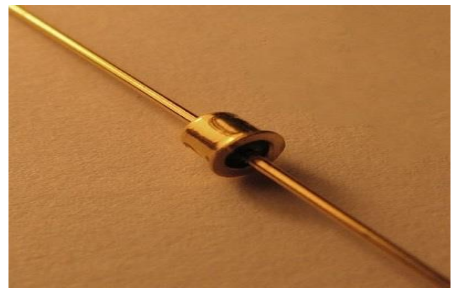
The Coupled-cavity TWT uses a slow wave structure of a series of cavities coupled to one another. The resonant cavities are coupled together with a transmission line. The electron beam (shown in Figure 9 as red beam) is velocity modulated by an RF input signal at the first resonant cavity. This RF energy (displayed as blue arrow) travels along the cavities and induces RF voltages in each subsequent cavity.

If the spacing of the cavities is correctly adjusted, the voltages at each cavity induced by the modulated beam are in phase and travel along the transmission line to the output, with an additive effect, so that the output power is much greater than the power input.

TUNNEL DIODE

The tunnel diode is a type of microwave semiconductor diode that can be used in oscillators and also amplifiers. Rather than using the standard physics of the ordinary PN junction, the tunnel diode uses a quantum mechanical effect called tunneling – from which it gains its name. The tunneling effect gives the tunnel diode a negative resistance region and this enables it to be used as an oscillator and also in pre-amplifier applications at frequencies well into the microwave region. They were used in television receiver front end oscillators and oscilloscope trigger circuits, etc.

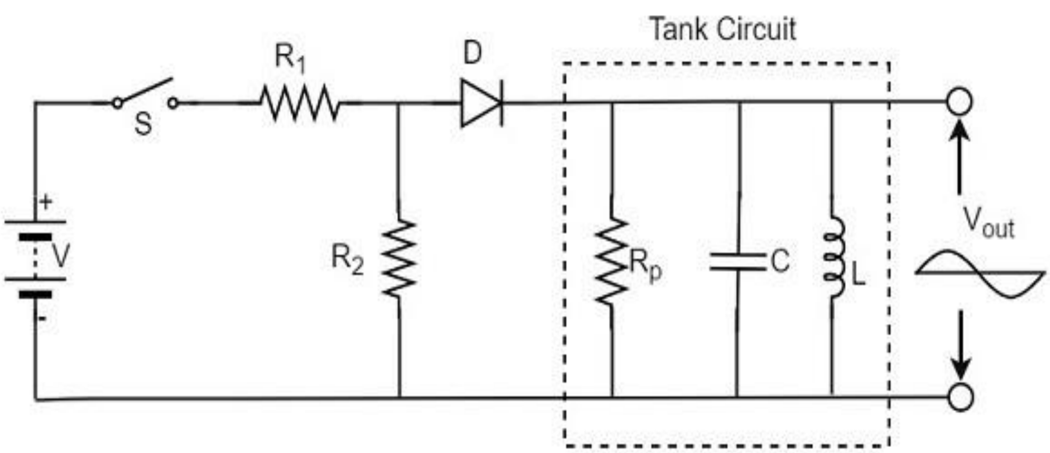




When the impurity concentration in a diode increases, the width of depletion region decreases, in turn extends some extra force to the charge carriers to cross the junction. When this concentration is further increased, due to less width of the depletion region and the increased energy of the charge carriers, they penetrate through the potential barrier, instead of climbing over it. This penetration can be understood as **Tunneling** and hence the name, **Tunnel diode**.

Tunnel Diode Oscillator

The tunnel diode helps in generating a very high frequency signal of nearly 10GHz. A practical tunnel diode circuit may consist of a switch S, a resistor R and a supply source V, connected to a tank circuit through a tunnel diode D.



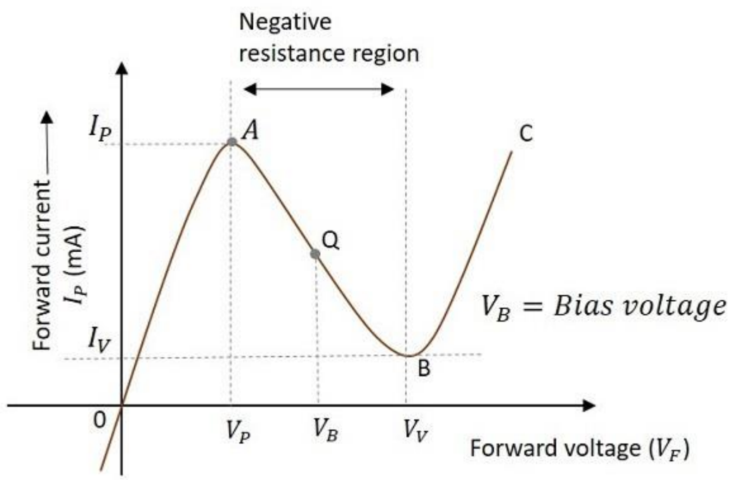
In this circuit, the resistor R_1 sets proper biasing for the diode and the resistor R_2 sets proper current level for the tank circuit. The parallel combination of resistor R_p inductor L and capacitor C form a tank circuit, which resonates at the selected frequency.

When the switch S is closed, the circuit current rises immediately towards the constant value, whose value is determined by the value of resistor R and the diode resistance. However, as the voltage drop across the tunnel diode V_D exceeds the peak-point voltage V_P, the tunnel diode is driven into negative resistance region.

In this region, the current starts decreasing, till the voltage V_D becomes equal to the valley point voltage V_V. At this point, a further increase in the voltage V_D drives the diode into positive resistance region. As a result of this, the circuit current tends to increase. This increase in circuit will increase the voltage drop across the resistor R which will reduce the voltage V_D.

V – I Characteristics

The following graph shows the V-I characteristics of a tunnel diode



The curve AB indicates the negative resistance region as the resistance decreases while the voltage increases. It is clear that the Q-point is set at the middle of the curve AB. The Q-point can move between the points A and B during the circuit operation. The point A is called **peak point** and the point B is called **valley point**.

During the operation, after reaching the point B, the increase in circuit current will increase the voltage drop across the resistor R which will reduce the voltage V_D. This brings the diode back into negative resistance region.

The decrease in voltage V_D is equal to the voltage V_P and this completes one cycle of operation. The continuation of these cycles produces continuous oscillations which give a sinusoidal output.

Advantages

The advantages of a tunnel diode oscillator are as follows

- It has high switching speeds.
- It can handle high frequencies.

Disadvantages

The disadvantages of a tunnel diode oscillator are as follows

- They are low power devices.
- Tunnel diodes are a bit costly.

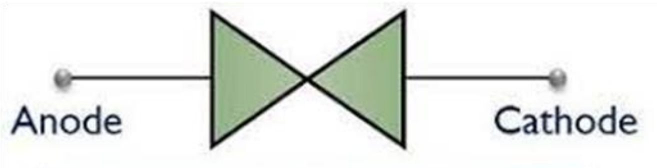
Applications

The applications of a tunnel diode oscillator are as follows

- It is used in relaxation oscillators.
- It is used in microwave oscillators.
- It is also used as Ultra high speed switching device.
- It is used as logic memory storage device.

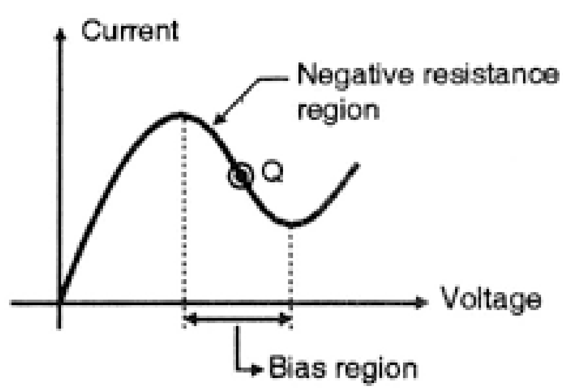
GUNN DIODES

Gunn diodes have been available for many years and they form a very effective method of generating microwave signals anywhere from around 1 GHz up to frequencies of possibly 100 GHz. Gunn diodes are also known as transferred electron devices, TED. Although is referred to as a diode, the devices does not possess a PN junction. Instead the device uses an effect known as the Gunn Effect (named after the discoverer, J B Gunn). Although the Gunn diode is normally used for generating microwave RF signals, the Gunn diode may also be used for an amplifier in what may be known as a transferred electron amplifier or TEA. As Gunn diodes are easy to use, they form a relatively low cost method for generating microwave RF signals, often being mounted within a waveguide to form a simple resonant cavity.



Symbol of Gunn Diode

The Gunn diode is a unique component - even though it is called a diode, it does not contain a PN diode junction. The Gunn diode or transferred electron device can be termed a diode because it has two electrodes. Gunn diode or Gunn device is just a slab of N-type gallium arsenide. Gallium arsenide has the interesting and unusual property that the mobility of the electrons actually decreases as the electric field strength increases over a certain range. Normally, larger the electric field strength, faster the movement of electrons through the conductor. With gallium arsenide and few other semiconductors such as indium arsenide and gallium phosphide, there is a region where the mobility actually decreases as the field increases.



VI Characteristics of Gunn Diode

The graph shown in the figure is said to have a negative resistance region. The differential resistance could be defined as

$$r = \frac{\Delta v}{\Delta i}$$

The differential resistance is negative over a certain range of field strength in gallium arsenide. A device with negative resistance can be made to work as an oscillator or as an amplifier. The negative resistance effect was discovered b J.B.Gunn in 1963. He found that when a voltage was applied across the gallium arsenide slab, oscillations occurred. The frequency of the oscillations depended on the thickness of the slab and had a period equal to the time required for an electron to pass through the slab at a drift velocity of $1x10^5$ m/s. The reason for the oscillation was that the domain with a large electric field formed in the material and moved toward the positive terminal. The domain forms when there is a small region with an electric field greater than in the rest of the material, due to some local impurity.

When this field reaches the threshold value, the electrons in the region move more slowly. A concentration of charge builds up, forming a domain which moves through the device, collecting most of the charges. When the domain reaches positive terminal, all of its electron leave at once, causing a pulse of current. Simultaneously, another domain forms and the process continues. This mode of operation is called the **transmit-time mode**. It can also be referred to as the **domain mode** or **Gunn mode**.

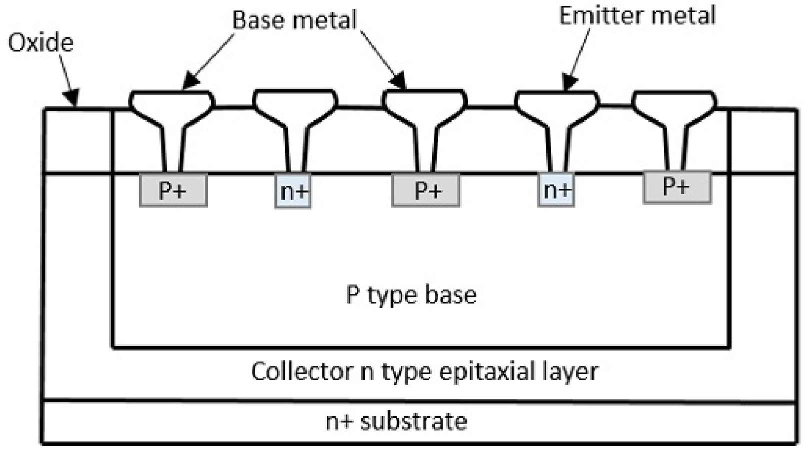
The Gunn mode provides oscillations at a frequency that depends on the device geometry and not on the external circuit. There are several other modes of operation that allow the device to be tuned by placing it in a resonant cavity. The effect of the cavity is to remove or quench, the domain before it reaches the anode terminal causing the device to operate at a higher frequency. The **limited space charge accumulation** (LSA) mode is the most common. In this mode, the device is biased in the negative resistance region, but the voltage swing in such that the device moves out of this region once per cycle, so that the domain is quenched.

The Gunn device can also be used as an amplifier, but like all two terminal negative resistance device, it requires a circulator to separate the input from the output.

MICROWAVE TRANSISTORS

There is a need to develop special transistors to tolerate the microwave frequencies. Hence for microwave applications, **silicon n-p-n transistors** that can provide adequate powers at microwave frequencies have been developed. They are with typically 5 watts at a frequency of 3GHz with a gain of 5dB.

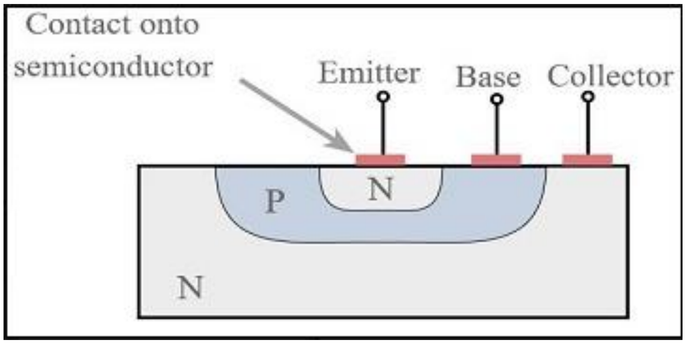
The micro wave bipolar transistor is a non linear device, which is mostly silicon npn type operating up to ‘5 GHz’. For high frequency applications, the NPN structure is preferred because the electron mobility is higher than hole mobility. Diffusion and ion implantation are the common methods used for transistor fabrication.



Construction

An epitaxial n layer is grown over a low resistivity n+ silicon substrate. Above the epitaxial layer, a p region is diffused forming the base and n+ layer is diffused over the p region to form the emitter. The silicon substrate acts as the collector.

The microwave bipolar transistors are active three terminal devices which is commonly used for amplification process and switching phenomena. The three regions of the transistor are emitter, base and collector. The emitter region forms the input of the device and the collector region forms the output of the device. The emitter region of the transistor is heavily doped and has moderate area of cross section. The base of the transistor is thin and lightly doped to reduce the recombination rate. The collector region of the transistor is large and moderately doped. The charge carriers from the emitter are supplied to the collector through the base. When the charge carriers from the emitter reach the base some of them recombine with the charge carriers in the base. The remaining charge carriers are directed towards the collector constituting the collector current or output current.

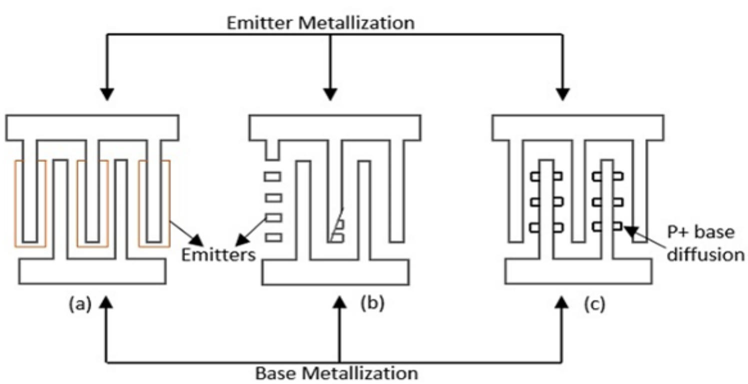


An **n** type epitaxial layer is grown on **n+** substrate that constitutes the collector. On this **n** region, a SiO_2 layer is grown thermally. A **p-base** and heavily doped **n-emitters** are diffused into the base. Connections are made in parallel.

Such transistors have a surface geometry categorized as either **interdigitated**, **overlay**, or **matrix**. These geometries have wide emitter area to overcome transit time limitations. The interdigitated geometry is used for small signal, small power circuits. The matrix geometry is sometimes called mesh or emitter grid. Overlay and Matrix structures are useful as power devices in the UHF and VHF regions.The over lay and matrix types are used for small power only. Power transistors employ all the three surface geometries.

Operation of Microwave Transistors

In a microwave transistor, initially the emitter-base and collector-base junctions are reverse biased. On the application of a microwave signal, the emitter-base junction becomes forward biased. If a **p-n-p** transistor is considered, the application of positive peak of signal forward biases the emitter-base junction, making the holes to drift to the thin negative base. The holes further accelerate to the negative terminal of the bias voltage between the collector and the base terminals. A load connected at the collector, receives a current pulse.



Surface geometry of n-p-n microwave transistor
(a) Inter-digitated (b) Overlay (c) Matrix

Power Frequency Limitations:

Microwave transistors have limitations on frequency and power. These limitations can be due to maximum velocity of carriers, maximum electric field and maximum current. The four basic equations for the power frequency limitation are

1. Voltage – Frequency limitation:

$$V_m f_T = \frac{E_m V_s}{2\pi}$$

Where, f_T - Cut off frequency $= \frac{1}{2\pi T}$

$T = {}^{L}/_{V}$

V_m - maximum voltage

V_s – velocity

E_m – Maximum electric field

When the length decreases, the average time T decreases. As a result, the frequency increases. When the frequency increases, the applied maximum voltage decreases.

2. Current frequency Limitation:

$$I_m X_C f_T = \frac{E_m V_s}{2\pi}$$

Where, X_C – Impedance

I_m – Maximum Current

If the impedance level is zero, the maximum current is infinite. The impedance value should be maintained in such a way, that maximum current is obtained for producing maximum power.

3. Power - frequency Limitation:

$$\sqrt{(P_m X_C)} f_T = \frac{E_m V_s}{2\pi}$$

P_m – Maximum Power

If the value of X_C is zero, the maximum power delivered is infinite.

4. Power Gain frequency Limitation:

$$\sqrt{(G_m V_m V_{th})} f_T = \frac{E_m V_s}{2\pi}$$

G_m – Maximum Gain

V_m – Maximum Voltage

V_{th} – Thermal Voltage

If the frequency increases, the gain of the device decreases.

HORN ANTENNAS

Horn antennas can be viewed as impedance transformers that match waveguide impedances to that of free space. The examples in the figure represent the most common types.

The E – plane and H – plane sectoral horns are named for the plane in which the horn flares; the pyramidal horn flares in both planes. The conical horn is the most appropriate with circular waveguide.

The gain and directivity of horn antennas depend on the type of horn and its dimensions. For pyramidal horn (the most common type), its gain is proportional to both the aperture dimensions d_E and d_H. The flare angle is limited by impedance matching considerations. So to attain high gain, a long unwieldy horn is required. The equation for gain is

$$G = \frac{7.5 d_E d_H}{\lambda^2}$$

d_E – E plane aperture

d_H – H plane aperture

The beam width is different in the two directions. In the H-plane, it is $\theta_H = \frac{70\lambda}{d_H}$ and in the E-plane, it is $\theta_E = \frac{56\lambda}{d_E}$. For practical horns, the gain is in the vicinity of 20dBi with a beam width of about 25°. It works over a frequency range of approximately 2:1.

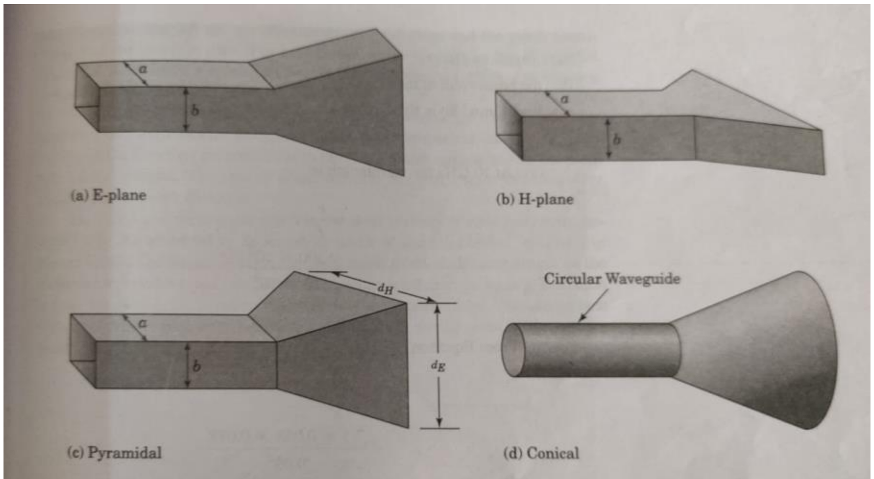


Figure 17.42 Horn antennas

TERMINAL AND REPEATER SITING

In general, a microwave system should use as repeaters as possible. Repeaters cost money and each one increases the chances of an equipment breakdown that can disable the link. More importantly additional links contribute to noise levels in analog systems and increase the jitter in digital ones. On the other hand, repeater stations must not be located beyond line of sight propagation range from each other and all sorts of practical considerations can prevent certain sites from being used for repeaters. Land must be acquired, access and electrical power must be arranged and the topography must be inspected for the best repeater sites, preferably on high points of the terrain.

Terrestrial microwave systems use relatively low power transmitters with high gain parabolic or hg horn antennas. By concentrating the transmitted power into narrow beam, these antennas increase the effective power and reduce interference to and from other systems. There are limits to antenna gain. Gain greater than 45dB should be avoided because antennas with high gain will have low beam width (less than one degree) such that mounting requirements are severe. Slight motions of the antenna tower due to wind can be more than enough to cause signal loss with such narrow antenna beams.

Feed lines between transmitters and/or receivers and antennas are always constructed at frequencies above 2 GHz, to reduce losses. At lower frequencies, coaxial cable may be used.

Terrestrial microwave links generally use line of sight (LOS) propagation. The maximum distance between two stations depends on the height of the transmitting and receiving antennas as well as on the nature of the terrain between them. For an average atmospheric conditions and level terrain, the distance over which LOS propagation is possible and is given by

$$\mathbf{d} = \sqrt{17\mathbf{h_T}} + \sqrt{17\mathbf{h_R}}$$

d – Maximum distance in kms

h_T - Height of the transmitting antenna in m

h_R – Height of the receiving antenna in m

The above equation gives approximate results only. Fr accurate calculations, it is necessary to obtain topographic maps of the area and note any obstacles in the path between transmitter and receiver. The path is then plotted on graph paper. There are two common methods of doing this. The first is to use ordinary graph paper and adjust the heights of the obstacles and the terrain to compensate for the curvature of the earth. The second way is to use special graph paper known as 4/3 earth paper, on which the horizontal lines are curved to represent the earth's curvature. For radio waves, the earth looks about one-third larger than it really is, hence the term 4/3 earth paper is used.

PATH CALCULATIONS

Once the path for a microwave link has been determined, it is necessary to ensure that the received signal power is sufficient for the required signal noise ratio. The variables involved here include the power of the transmitter, the receiver noise figure and the gain of their respective antennas.

The power density of a signal under free space propagation conditions is given by

$$\mathbf{P_D = \frac{EIRP}{4\pi r^2}} \qquad \text{---- (1)}$$

P_D -Power density

EIRP –Effective isotropic radiated power

r –Distance from antenna

A more known version of the same equation is

$$\frac{P_R}{P_T} = G_T(dBi) + G_R(dBi) - (32.44 + 20\log d + 20\log f) \qquad \text{-- (2)}$$

Carrier to Noise Ratio

For analog microwave systems, satisfactory performance is normally defined as the carrier to noise ratio that exceeds a given number of decibels. The carrier to noise ratio is simply the signal to noise ratio measured before the signal is demodulated.

In order to determine the carrier to noise ratio, we need the signal power and the noise power. The signal power can be found using equation (2), while noise mainly consists of thermal noise either received by the antenna, or generated in the antenna, transmission line or receiver. The easiest way to combine these sources is to find the corresponding noise temperature of each, referenced to the receiver input, add the noise temperatures and then find the noise power. The noise power is given by

$$\mathbf{P_N = kTB}$$

k – Boltzmann's constant

T – Absolute temperature

B – Noise power bandwidth

The antenna receives noise from the sky and from the earth, depending on whether the beam of the antenna includes the ground. The noise temperature of the sky depends on the angle of elevation of the antenna, the frequency and atmospheric conditions. Resistive losses in the antenna and its feed line must also be taken into account to get an equivalent noise temperature at the receiver input. The equivalent noise temperature at the receiver input is given by

$$T_{eq} = \frac{(L-1)T_R + T_{sky}}{L}$$

T_{eq} – Effective noise temperature of antenna and feed line

L - Loss in feed line and antenna as a ratio of input to output power

T_R – Reference Temperature

T_{sky} – Effective sky temperature

For a reference temperature of 290K, the equation becomes

$$T_{eq} = \frac{(L-1)290 + T_{sky}}{L}$$

In some cases, equivalent noise temperature at the receiver input has been found. In such cases we can add the noise temperature of the receiver. Sometimes receiver noise temperature is specified, but is given as noise figure. In such cases, it is easy to convert noise figure to noise temperature

$T_{eq} = 290(NF - 1)$ where, NF – Noise figure

Once the noise temperature of the antenna-feedline combination and the receiver has been determined, they can be summed to find the noise temperature of the system. From that information and the system bandwidth, it is easy to determine the noise power. The carrier to noise ratio can be found using the signal power at the antenna input

Energy per Bit/ Noise Density Ratio

In digital communication schemes, it is common to specify the noise performance of a system in terms of the ratio of the energy per received information bit to the noise density.

The noise density is the noise power in one hertz of the spectrum. The energy per bit per noise density is directly related to the bit error rate for the system. The energy per bit is the energy received in the time taken to transmit one bit, i.e. it is the received signal power multiplied by the period for one bit.

Since period is the inverse of frequency, it is easy to calculate the energy per bit, if the received signal power and the bit rate are given:

$$E_b = \frac{P_R}{f_b}$$

E_b – Energy per in joules

P_R – Received signal power in Watts

F_b – Bit rate in bit per second.

The noise power density is given by $N_O = kT$

N_O – Noise power density in Watts per Hertz

T – Temperature in Kelvin

k – Bolzmann Constant

Required E_b/N_o ratio's for a bit error rate of 1 x10^{-4} vary from about 10 to 20 dB for practical digital modulation schemes.

Fading

Any microwave system must include an allowance for fading in its system gain calculations. Fading is a reduction in signal strength below its nominal level. It has a number of causes, including

- "Multipath reception" – in which a direct signal is partially cancelled by reflections from ground or water
- "Attenuation due to rain" – mainly at frequencies above 10 GHz
- "Ducting" – in which signals are deflected by layers of different temperature and humidity in the atmosphere
- "Aging or partial failure" – in transmitting and receiving equipments

There are two basic methods for dealing with fading. One is to overbuild the system to make sure that the system has sufficient gain to achieve the desired signal to noise ratio or bit error rate, even when fading takes place. The additional gain called fade margin can be obtained by increasing transmitter power, antenna gain or receiver sensitivity. Another approach called diversity can reduce fading without restoring to extreme power levels. As fading is frequency selective, changing the frequency slightly can eliminate the problem.

Let us consider the signal reaches the receiving antenna by two paths, one direct and the other reflected from water. Fading occurs when the path lengths has a phase difference between the two signal is 180°. The variations in atmospheric conditions can easily change the path length of microwave wavelengths to change the interference from constructive back to destructive.

Fading due to multipath propagation can be avoided by slightly changing the frequency so that the phase difference is n longer 180º. This technique is called frequency diversity.

To protect against fading on a moment to moment basis, frequency diversity requires two transmitters and two receivers, separated in frequencies. Providing dual transmitters and receivers is expensive, but it allows hot standby protection. If one transmitter or receiver should fail, communication can continue uninterrupted.

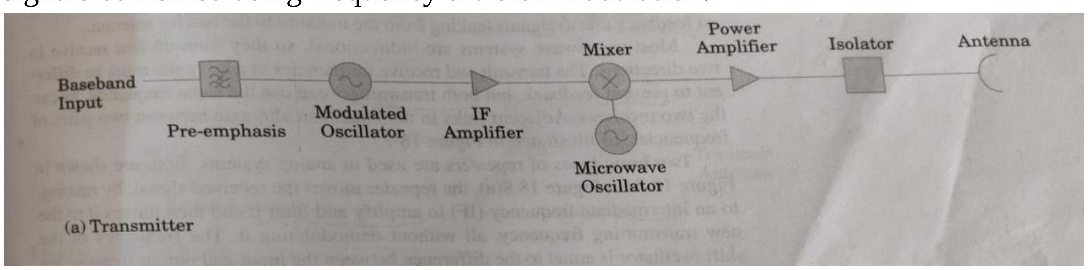
Another way to prevent multipath fading is to change the path length by moving either the transmitting or the receiving antenna. This technique is called **space diversity**, generally involves placing two antennas one above the other on the same tower. The two antennas should be separated by 200 wavelengths or more. Space diversity requires taller towers as well as more antennas.

FIXED MICROWAVE LINKS

Terrestrial microwave links can be either analog or digital systems. Analog systems use either frequency modulation (FM) or single sideband suppressed carrier amplitude modulation (SSB). Digital systems use phase –shift keying or quadrature amplitude modulation (QAM).

FM systems

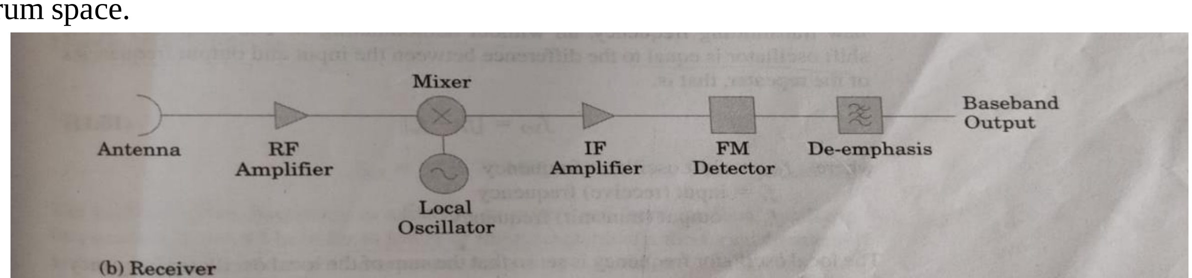
Typically the IF is 70 MHz and narrowband FM is used, so that the transmitted bandwidth is only slightly more than twice that of the baseband signal. The baseband signal shown at the transmitter input and the receiver output can consist of a single broadcast quality television signal or a number of telephone signals combined using frequency division modulation.



(a) Transmitter

The transmitter block diagram is shown in figure. The FM signal is generated at a relatively low frequency, then mixed up to the required transmit carrier frequency. The local oscillator can be a microwave oscillator such as klystron or Gunn diode. A TWT is commonly used in the output stage or a solid state power amplifier using FETs can be used.

The receiver block diagram is shown below. A RF amplifier is placed before the mixer in modern designs. Nowadays, gallium arsenide FET transistor in a low noise amplifier is used before the mixer. The intermediate frequency is generally 70 MHz, but the IF bandwidth depends on the bandwidth of the signal. More than one carrier is used depending upon the traffic capacity and the available spectrum space.



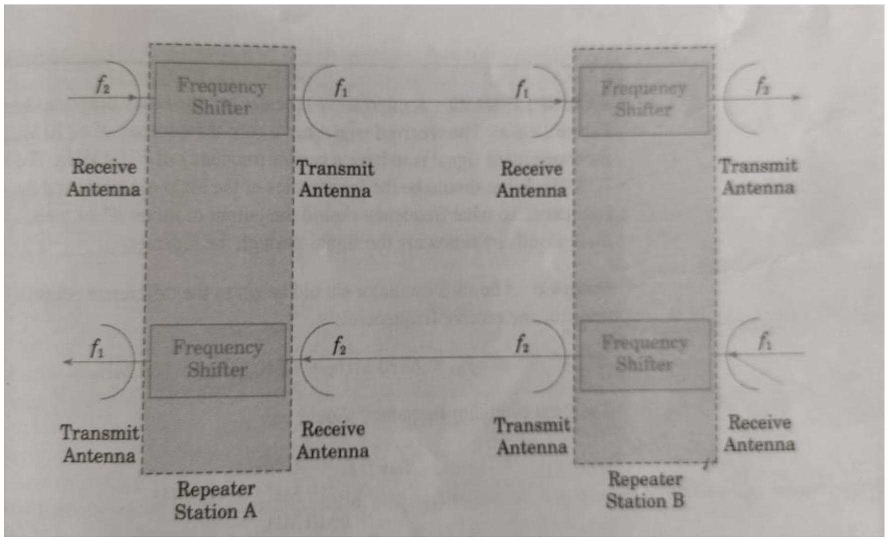
(b) Receiver

Digital data can also be transmitted with FM systems using external modems. When new systems are designed for digital signals, it is more efficient to use digital modulation schemes. When a microwave system extends the LOS distance, repeaters are necessary. A repeater receives from one direction on one frequency and simultaneously transmits in a different direction on another frequency. Different frequencies must be used for transmission and reception to avoid the possibility of feedback due to signals leaking from transmission to the receiving antenna.

Most of the microwave systems are bidirectional, so they transmit and receive in two directions. The transmitter and receiver frequencies at a relay site must be different to prevent feedback, but both transmitters can use the same frequency, as can the two receivers.

Two basic types of repeaters are used in analog systems: IF repeater and baseband repeater

IF Repeater



The IF repeater moves the received signal by mixing to an intermediate frequency to amplify and filter it and then moves it to the new transmitter frequency, all without demodulating it. The frequency of the shift oscillator is equal to the difference between the input and output frequencies of the repeater.

i.e. $\boldsymbol{f_{so}} = |\boldsymbol{f_i} - \boldsymbol{f_o}|$

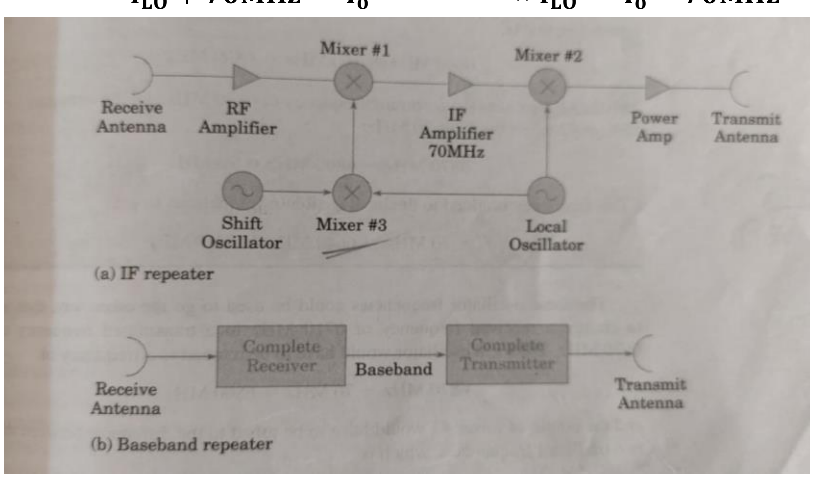
where f_{so} – Shift oscillator frequency

f_i – Input (Receive) frequency

f_o – Output (transmit) frequency

The local oscillator frequency is set so that the sum of the local oscillator frequency and the intermediate frequency is equal to the output carrier frequency.

$$\mathbf{f_{LO} + 70MHz = f_o} \qquad \therefore \mathbf{f_{LO} = f_o - 70MHz}$$



The shift oscillator frequency is added to or subtracted from the local oscillator frequency in mixer #3 in order to respectively lower or raise the signal frequency as the signal moves through the repeater.

In a baseband repeater, the signal is demodulated, then re-transmitted. This type of repeater is just a receiver with its output connected to the modulation input of the transmitter.

Single Sideband System (SSB)

Some recent analog microwave systems use SSB in order to reduce the bandwidth required. The bandwidth required for SSB transmission is equal to the baseband bandwidth while the bandwidth needed for FM is always more than twice as the baseband bandwidth. This bandwidth conserving property of SSB allows a very considerable increase in traffic handling capacity along busy routes. It is interesting to note that analog voice channels are multiplexed using SSB before they reach the microwave transmitter, whether the transmitter itself uses FM or SSB.

SSB microwave equipment resembles that used for FM, except that the IF is usually 74.13 MHz instead of 70MHz and great care must be taken to preserve linearity in all amplifiers. Like any amplitude modulation system, SSB requires linear amplification if distortion of the baseband signal is to be avoided. Most SSB microwave transmitters use TWT for the transmitter power amplifier stage like FM systems. SSB radio links can be used to transmit digital signals and this will become more important as the telephone system gradually converts to digital operation.

Microwave Digital Radio

Digital techniques have two main advantages in microwave radio links.

First, the accumulation of noise as the signal travels through many links can be avoided by regenerating the signal at each repeater. This requires demodulating the signal, decoding the data, then recoding and re-modulating on a new carrier. As long as the signal to noise ratio for each single link is high enough to avoid errors, there will be no increase in error rates as the signal progresses through the system. In practical, a zero error rate is not obtainable, but the error rates will add from link to link, rather than multiplying as in analog systems.

The second major reason for employing digital modulation techniques in microwave radio is to maintain compatibility with digitally coded baseband signals. Digital signals using time division multiplexing can pass through digital microwave systems without alteration. The only disadvantage of digital transmission of voice and video signals is that it requires more bandwidth than analog FM or SSB.

To be transmitted by microwaves, a digital signal must modulate a microwave frequency carrier. Two modulation schemes are used for digital microwave radio: phase shift keying (PSK) and quadrature amplitude modulation (QAM), which involves both amplitude and phase shifts. Microwave systems have very much higher carrier frequencies, bandwidths, and data rates, but the underlying principles are the same.

In general, the more recent and the more sophisticated the system, the more bits per symbol is used. The number of bits of information transmitted with each symbol is a function of the number of possible states:

$$\boldsymbol{N_B = \log_2 N_S}$$

N_S – Number of possible states per symbol

Transmitting more bits per symbol yields a higher data rate for a given RF bandwidth. On the other hand, transmitting more information per symbol requires greater precision at both transmitter and receiver, so that small amplitude and phase changes can be generated and detected reliably. A better signal to noise ratio is also required for a given error rate with more bits per symbol.

Most digital radio systems use 16-QAM or 64-QAM. The 16-QAM uses four amplitudes and four phase angles four a total of 16 possible states or 4 bits per symbol. The 64-QAM has eight amplitudes and eight phases, giving 64 states or 6 bits per symbol. The current state of art is 256-QAM which transmits 8 bits per symbol. Experimental use has been made of 1024-QAM. All QAM schemes have the disadvantage, common to all AM systems, of requiring linear amplifiers in the transmitter. FSK and PSK schemes do not require this linearity, so the transmitters can be more efficient. On the other hand, QAM has greater noise immunity than FSK or PSK for a given transmitter power and data rate and it requires less bandwidth.

Aside from the modulator, transmitters and receivers for digital microwave radio resemble those for analog systems. Repeaters must demodulate the signal to baseband to achieve the advantages of digital transmission, i.e. they must be regenerative repeaters. Such repeaters can restore a signal that has been distorted or corrupted by noise to its original state, as long as the signal is readable at the receiver.

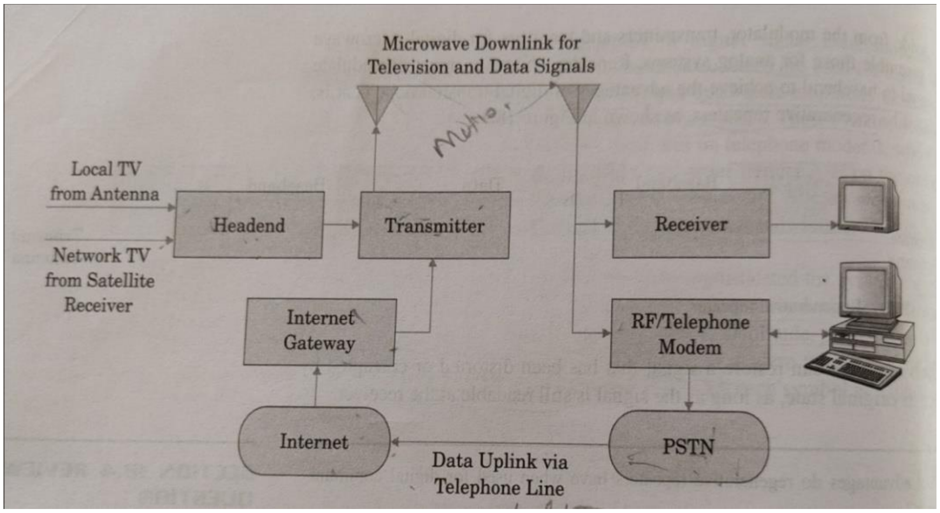
LOCAL MICROWAVE DISTRIBUTION SYSTEMS

Local microwave or multipoint distribution systems (LMDS) use terrestrial microwave transmission to provide a number of services to homes and businesses. These services include broadcast television, high speed internet access, and telephony.

These are also known as local multipoint communication systems (LCMS). Most systems use frequencies in the 28-GHz range. There is also a variant known as multichannel multipoint distribution system (MMDS) which operates in the 2-GHz range.

MMDS

MMDS have been operating in the 2-GHz range. The earliest systems were analog, but recent versions were digital. MMDS began as an alternative to cable television using coaxial cable and is sometimes called by the contradictory name **wireless cable**.

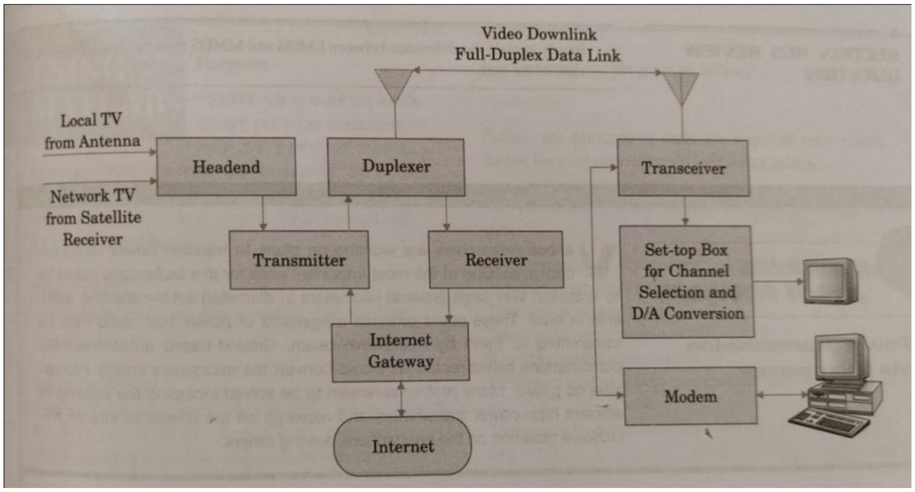


A typical MMDS is shown in the figure below. Local television stations are received off the air at the headend, which also picks up cable-network stations by satellite. The signals are rebroadcast in a digital format called MPEG (Motion Picture Experts Group) by microwave from one or more tall towers to small wall mounted antennas on houses and apartment buildings. The system works over line of sight distances of up tp about 50 km, depending on the antenna heights. Receiving antennas must have a direct line of sight to the transmitter, so the system is not suitable in hilly areas or dense metropolitan areas with tall buildings.

MMDS are also used for Internet access. All data requests from all subscribers are transmitted on one RF channel, and each terminal accepts only those packets that are addresses to it. In this way, the system resembles a LAN. MMDS are one way system. A bidirectional system would require dividing the coverage area into smaller cells, which makes it expensive.

With an existing MMDS, any return link that may be required for uploading data to the internet or for ordering pay per view television programs must use some other technology. Recently, MMDS operators have begun to investigate the possibility of using a low speed wireless link, such as PCS (Personal Communication Services) for upstream communication. The only problem with this is that its cost.

LMDS



The concept behind an LMDS is similar to that behind the earlier MMDS, but with some major changes.

First, moving the frequency to 28 GHz allows much freer use of spectrum. Second, the high frequency causes severe attenuation problems in the presence of rain, and even foliage from trees can block the signal. LMDS is designed to be bidirectional. This allows it to be used for internet access without using the PSTN (Public Switched telephone Network) and in fact, allows the LMDS to be used for telephone communication instead of PSTN. However, early versions of LMDS still rely o n a wire line modem and a PSTN connection for upstream traffic.

Because of the very large amount of spectrum available, LMDS can offer very high data rates. As with MMDS, LMDS is essentially a line of sight technology. Due to the high frequency and short wavelength, LMDS may be able to take advantage of reflections from buildings to achieve coverage in some areas that are not in the direct line of sight from a transmitter.

Summary

- The transmitters in communication satellites and their ground states often use TWT amplifiers. TWT is a linear beam tube. It has a slow wave structure. It may consist of a series of coupled cavities, but it has the form of a helix.
- The tunnel diode is a type of microwave semiconductor diode that can be used in oscillators and also amplifiers. They were used in television receiver front end oscillators and oscilloscope trigger circuits, etc.
- Gunn diodes have been available for many years and they form a very effective method of generating microwave signals anywhere from around 1 GHz up to frequencies of possibly 100 GHz. Gunn diodes are also known as transferred electron devices, TED.

- ✶ The carrier to noise ratio is simply the signal to noise ratio measured before the signal is demodulated.
- ✶ The energy per bit is the energy received in the time taken to transmit one bit, i.e. it is the received signal power multiplied by the period for one bit.
- ✶ Terrestrial microwave links can be either analog or digital systems. Analog systems use either frequency modulation (FM) or single sideband suppressed carrier amplitude modulation (SSB). Digital systems use phase –shift keying or quadrature amplitude modulation (QAM).
- ✶ Local microwave or multipoint distribution systems (LMDS) use terrestrial microwave transmission to provide a number of services to homes and businesses. These services include broadcast television, high speed internet access, and telephony.

Review Questions

Two Marks

1. What are microwaves?
2. Give the applications of microwaves.
3. What is the importance of microwaves?
4. Write down the components of TWT.
5. Give the different types of slow wave structures.
6. What is a tunnel diode?
7. What are the advantages of tunnel diode?
8. Give the applications of tunnel diode.
9. Define Gunn diode.
10. State Gunn Effect.
11. Write down the applications of Gunn diode.
12. Write short note on path calculations.
13. What is meant by surface acoustic wave?
14. What is baseband repeater?
15. Define carrier to noise ratio.

Five Marks

1. Briefly explain the tunnel diode.
2. Discuss in detail about the Gunn diode.
3. Explain microwave transistors.
4. Explain Horn antenna.
5. Briefly explain the local microwave distribution system.

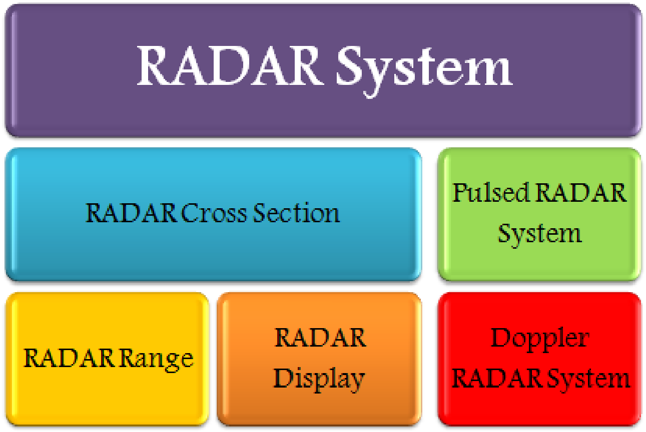
Ten Marks

1. Briefly explain the construction and working of travelling wave tube.
2. Explain the concept of path calculations in microwave propagation.
3. Explain fixed microwave links.

Unit V

Radar System

RADAR Concept – RADAR cross section – RADAR Range – RADAR Display – RADAR system Circuit and components – Pulsed RADAR System – Doppler RADAR system.



Introduction

RADAR stands for Radio Detection and Ranging System. It is basically an electromagnetic system used to detect the location and distance of an object from the point where the RADAR is placed. It works by radiating energy into space and monitoring the echo or reflected signal from the objects. It operates in the UHF and microwave range.

Radar is an electromagnetic sensor, used to notice, track, locate, and identify different objects which are at certain distances. The working of radar is, it transmits electromagnetic energy in the direction of targets to observe the echoes and returns from them. Here the targets are nothing but ships, aircraft, astronomical bodies, automotive vehicles, spacecraft, rain, birds, insects, etc. Instead of noticing the target's location and velocity, it also obtains their shape and size sometimes.

The main objective of radar as compared with infrared and optical sensing devices is to discover faraway targets under difficult climate conditions & determines their distance, range, through precision. Radar has its own transmitter which is known as a source of illumination for placing targets. Generally, it works in the microwave area of the electromagnetic spectrum that is calculated in hertz when frequencies extend from 400 MHz to 40 GHz. The essential components which are used in the radar

Radar undergoes quick development during the years 1930-the 40s to reach the requirements of the military. It is still broadly used through the armed forces, wherever several technological advances have created. Simultaneously, radar is also utilized in civilian applications particularly in controlling air traffic, observation of weather, navigation of ship, environment, sensing from remote areas, observation of planetary, measurement of speed in industrial applications, space surveillance, law enforcement, etc.

Working Principle

The **radar working principle** is very simple because it transmits electromagnetic power as well as examines the energy returned back to the target. If the returned signals are received again at the position of their source, then an obstacle is in the transmission way. This is the working principle of radar.

Fundamentals of Radar

The RADAR system generally consists of a transmitter that produces an electromagnetic signal which is radiated into space by an antenna. When this signal strikes an object, it gets reflected or reradiated in many directions. This reflected or echo signal is received by the radar antenna which delivers it to the receiver, where it is processed to determine the geographical statistics of the object.

The range is determined by calculating the time taken by the signal to travel from the RADAR to the target and back. The target's location is measured in angle, from the direction of the maximum amplitude echo signal, the antenna points to. To measure the range and location of moving objects, the Doppler Effect is used.

The essential parts of this system include the following.

- **A Transmitter:** It can be a power amplifier like a Klystron, Travelling Wave Tube, or a power Oscillator like a Magnetron. The signal is first generated using a waveform generator and then amplified in the power amplifier.
- **Waveguides:** The waveguides are transmission lines for transmission of the RADAR signals.
- **Antenna:** The antenna used can be a parabolic reflector, planar arrays, or electronically steered phased arrays.
- **Duplexer:** A duplexer allows the antenna to be used as a transmitter or a receiver. It can be a gaseous device that would produce a short circuit at the input to the receiver when the transmitter is working.
- **Receiver:** It can be a super heterodyne receiver or any other receiver which consists of a processor to process the signal and detect it.
- **Threshold Decision:** The output of the receiver is compared with a threshold to detect the presence of any object. If the output is below any threshold, the presence of noise is assumed.

How Does Radar use Radio?

Once the radar is placed on a ship or plane, then it requires a similar essential set of components to produce radio signals, transmit them into space and receive them by something, and finally display the information to understand it. A magnetron is one kind of device, used to generate radio signals which are used through radio. These signals are similar to light signals because they travel at the same speed but their signals are much longer with fewer frequencies.

The light signals wavelength is 500 nanometers, whereas the radio signals used by radar normally range from centimeters to meters. In an electromagnetic spectrum, both the signals like radio and light are made with variable designs of magnetic and electrical energy throughout the air. The magnetron in radar generates microwaves the same as a microwave oven. The main disparity is that the magnetron within radar has to transmit the signals several miles, rather than just small distances, so it is more powerful as well as much larger.

Whenever the radio signals have been transmitted, then an antenna functions as a transmitter to transmit them into the air. Generally, the antenna shape is bent so it mainly focuses the signals into an exact and narrow signal; however radar antennas also normally revolve so they can notice actions over a huge area.

The radio signals travel outside from the antenna with 300,000 km per second speed until they strike something and some of them return back to the antenna. In a radar system, there is an essential device namely a duplexer. This device is used to make the antenna change from side to side in between a transmitter & a receiver.

Types of Radar

There are different types of radars which include the following.

Bi-static Radar

This type of radar system includes a Tx-transmitter & an Rx- receiver that is divided through a distance that is equivalent to the distance of the estimated object. The transmitter & the receiver are situated at a similar position is called a monastic radar whereas the very long-range surface to air & air to air military hardware uses the bi-static radar.

Doppler Radar

It is a special type of radar that uses the Doppler Effect to generate data velocity regarding a target at a particular distance. This can be obtained by transmitting electromagnetic signals in the direction of an object so that it analyzes how the action of the object has affected the returned signal's frequency. This change will give very precise measurements for the radial component of an object's velocity within relation toward the radar. The applications of these radars involve different industries like meteorology, aviation, healthcare, etc.

Mono-pulse Radar

This kind of radar system compares the obtained signal using a particular radar pulse next to it by contrasting the signal as observed in numerous directions otherwise polarizations. The most frequent type of mono-pulse radar is the conical scanning radar. This kind of radar evaluates the return from two ways to measure the position of the object directly. It is significant to note that the radars which are developed in the year 1960 are mono-pulse radars.

Passive Radar

This kind of radar is mainly designed to notice as well as follow the targets through processing indications from illumination within the surroundings. These sources comprise communication signals as well as commercial broadcasts. The categorization of this radar can be done in the same category of bi-static radar.

Instrumentation Radar

These are designed for testing aircraft, missiles, rockets, etc. They give different information including space, position & time both in the analysis of post-processing & real-time.

Weather Radars

These are used to detect the direction and weather by using radio signals through circular or horizontal polarization. The frequency choice of weather radar mainly depends on a compromise of performance among attenuation as well as precipitation refection as an outcome of atmospheric water steam. Some types of radars are mainly designed to employ Doppler shifts to calculate the wind speed as well as dual-polarization to recognize the types of rainfall.

Mapping Radar

These radars are mainly used to examine a large geographical area for the applications of remote sensing & geography. As a result of synthetic aperture radar, these are restricted to quite stationary targets. There is some particular radar systems used to detect humans after walls that are more different as compared with the ones found within construction materials.

Navigational Radars

Generally, these are the same to search radars but, they available with small wavelengths that are capable of replicating from the ground & from stones. These are commonly used on commercial ships as well as long-distance airplanes. There are different navigational radars like marine radars which are placed commonly on ships to avoid a collision as well as navigational purposes.

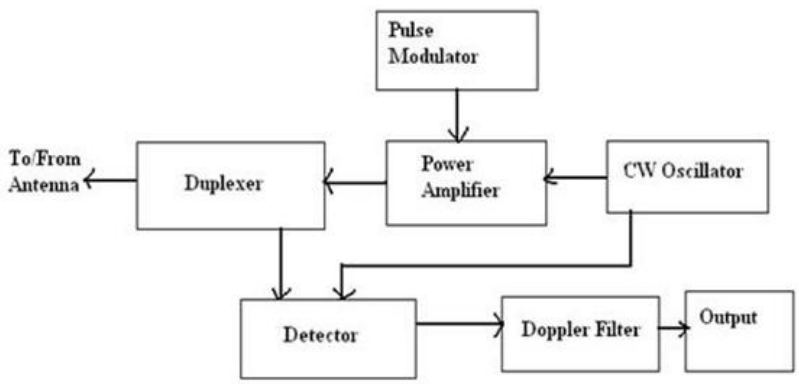
Pulsed RADAR

Pulsed RADAR sends high power and high-frequency pulses towards the target object. It then waits for the echo signal from the object before another pulse is sent. The range and resolution of the RADAR depend on the pulse repetition frequency. It uses the Doppler shift method.

The principle of RADAR detecting moving objects using the Doppler shift works on the fact that echo signals from stationary objects are in the same phase and hence get canceled while echo signals from moving objects will have some changes in phase. These radars are classified into two types.

Pulse-Doppler

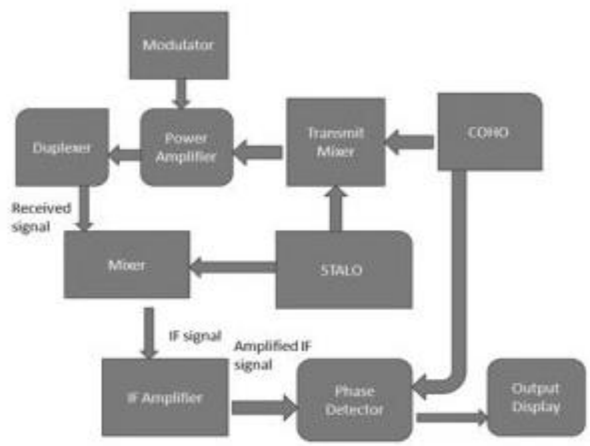
It transmits high pulse repetition frequency to avoid Doppler ambiguities. The transmitted signal and the received echo signal are mixed in a detector to get the Doppler shift and the difference signal is filtered using a Doppler filter where the unwanted noise signals are rejected.



Moving Target Indicator

It transmits low pulse repetition frequency to avoid range ambiguities. In an MTI RADAR system, the received echo signals from the object are directed towards the mixer, where they are mixed with the signal from a stable local oscillator (STALO) to produce the IF signal.

This IF signal is amplified and then given to the phase detector where its phase is compared with the phase of the signal from the Coherent Oscillator (COHO) and the difference signal is produced. The Coherent signal has the same phase as the transmitter signal. The coherent signal and the STALO signal are mixed and given to the power amplifier which is switched on and off using the pulse modulator.



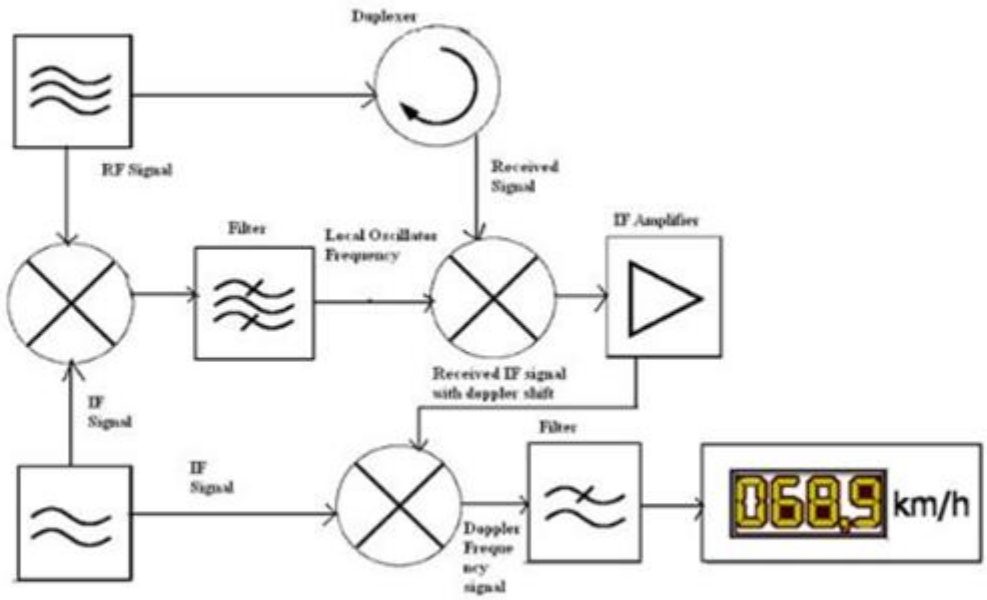
MTI Radar

Continuous Wave

The continuous wave RADAR doesn't measure the range of the target but rather the rate of change of range by measuring the Doppler shift of the return signal. In a CW RADAR electromagnetic radiation is emitted instead of pulses. It is basically used for speed measurement.

The RF signal and the IF signal are mixed in the mixer stage to generate the local oscillator frequency. The RF signal is then transmitted signal and the received signal by the RADAR antenna consist of the RF frequency plus the Doppler shift frequency. The received signal is mixed with the local oscillator frequency in the second mixture stage to generate the IF frequency signal.

This signal is amplified and given to the third mixture stage where it is mixed with the IF signal to get the signal with Doppler frequency. This Doppler frequency or Doppler shift gives the rate of change of range of the target and thus the velocity of the target is measured.



Block Diagram Showing CW RADAR

Radar Range Equation

The distance between Radar and target is called **Range** of the target or simply range, R. We know that Radar transmits a signal to the target and accordingly the target sends an echo signal to the Radar with the speed of light, C.

Let the time taken for the signal to travel from Radar to target and back to Radar be 'T'. The two way distance between the Radar and target will be 2R, since the distance between the Radar and the target is R.

Now, the following is the formula for **Speed**.

$$\mathbf{Speed} = \frac{\mathbf{Distance}}{\mathbf{Time}}$$

$$\mathbf{Distance} = \mathbf{Speed} \times \mathbf{Time}$$

$$\mathbf{2R} = \mathbf{C} \times \mathbf{T}$$

$$\mathbf{R} = \frac{\mathbf{CT}}{\mathbf{2}}$$

Radar range equation is useful to know the range of the target **theoretically**. There are two modified forms of Radar range equation. We will get those modified forms of Radar range equation from the standard form of Radar range equation. Now, let us discuss about the derivation of the standard form of Radar range equation.

Derivation of Radar Range Equation

The standard form of Radar range equation is also called as simple form of Radar range equation. Now, let us derive the standard form of Radar range equation. We know that **power density** is nothing but the ratio of power and area. So, the power density, P_{di} at a distance, R from the Radar can be mathematically represented as

$$\mathbf{P_{di} = \frac{P_t}{4\pi R^2}}$$

Where, P_t is the amount of power transmitted by the Radar transmitter

The above power density is valid for an isotropic Antenna. In general, Radars use directional Antennas. Therefore, the power density, P_{dd} due to directional Antenna will be

$$\mathbf{P_{dd} = \frac{P_t G}{4\pi R^2}}$$

Target radiates the power in different directions from the received input power. The amount of power, which is reflected back towards the Radar depends on its cross section. So, the power density P_{de} of echo signal at Radar can be mathematically represented as

$$\mathbf{P_{de} = P_{dd}\left(\frac{\sigma}{4\pi R^2}\right)}$$

Substitute, P_{dd} in P_{de}.

$$\mathbf{P_{de} = \left(\frac{P_t G}{4\pi R^2}\right)\left(\frac{\sigma}{4\pi R^2}\right)}$$

The amount of **power,** P_r **received** by the Radar depends on the effective aperture, A_e of the receiving Antenna.

$$\mathbf{P_r = P_{de} A_e}$$

Substituting P_{de} in P_r.

$$\mathbf{P_r = \left(\frac{P_t G}{4\pi R^2}\right)\left(\frac{\sigma}{4\pi R^2}\right) A_e}$$

$$\mathbf{P_r = \frac{P_t G \sigma A_e}{(4\pi)^2 R^4}}$$

$$\mathbf{R^4 = \frac{P_t G \sigma A_e}{(4\pi)^2 P_r}}$$

$$\mathbf{R = \left(\frac{P_t G \sigma A_e}{(4\pi)^2 P_r}\right)^{1/4}}$$

Standard Form of Radar Range Equation

If the echo signal is having the power less than the power of the minimum detectable signal, then Radar cannot detect the target since it is beyond the maximum limit of the Radar's range. Therefore, we can say that the range of the target is said to be maximum range when the received echo signal is having the power equal to that of minimum detectable signal. We will get the following equation, by substituting $R=R_{Max}$ and $P_r=S_{min}$ in the above equation.

$$\mathbf{R_{max} = \left(\frac{P_t G \sigma A_e}{(4\pi)^2 S_{min}}\right)^{1/4}}$$

This equation represents the **standard form** of Radar range equation. By using the above equation, we can find the maximum range of the target.

Modified Forms of Radar Range Equation

We know the following relation between the Gain of directional Antenna, G and effective aperture, A_e.

$$\mathbf{G = \frac{4\pi A_e}{\lambda^2}}$$

Substituting the equation for G in R_{max}, we get

$$\mathbf{R_{max}} = \left[\frac{\mathbf{P_t \sigma A_e}}{\mathbf{(4\pi)^2 S_{min}}}\left(\frac{\mathbf{4\pi A_e}}{\mathbf{\lambda^2}}\right)\right]^{1/4}$$

$$\mathbf{R_{max}} = \left[\frac{\mathbf{P_t \sigma {A_e}^2}}{\mathbf{4\pi\lambda^2 S_{min}}}\right]^{1/4}$$

This equation represents the **modified form** of Radar range equation. By using the above equation, we can find the maximum range of the target.

We will get the following relation between effective aperture, A_e and the Gain of directional Antenna, G from

$$\mathbf{G} = \frac{\mathbf{4\pi A_e}}{\mathbf{\lambda^2}}$$

$$\mathbf{A_e} = \frac{\mathbf{G\lambda^2}}{\mathbf{4\pi}}$$

Substituting the expression for A_e in R_{max}, we get

$$\mathbf{R_{max}} = \left[\frac{\mathbf{P_t \sigma G}}{\mathbf{(4\pi)^2 S_{min}}}\left(\frac{\mathbf{G\lambda^2}}{\mathbf{4\pi}}\right)\right]^{1/4}$$

$$\mathbf{R_{max}} = \left[\frac{\mathbf{P_t \sigma G^2 \lambda^2}}{\mathbf{(4\pi)^3 S_{min}}}\right]^{1/4}$$

The above equation represents **another modified form** of Radar range equation. By using the above equation, we can find the maximum range of the target.

Note − Based on the given data, we can find the maximum range of the target by using one of these three equations namely

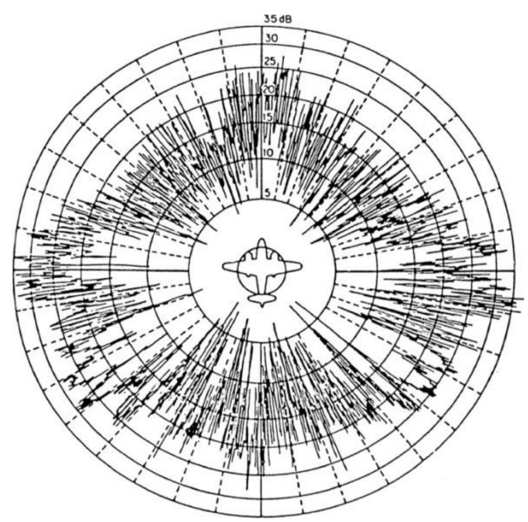
$$\mathbf{R_{max}} = \left[\frac{\mathbf{P_t G \sigma A_e}}{\mathbf{(4\pi)^2 S_{min}}}\right]^{1/4}$$

$$\mathbf{R_{max}} = \left[\frac{\mathbf{P_t \sigma {A_e}^2}}{\mathbf{4\pi\lambda^2 S_{min}}}\right]^{1/4}$$

$$\mathbf{R_{max}} = \left[\frac{\mathbf{P_t \sigma G^2 \lambda^2}}{\mathbf{(4\pi)^3 S_{min}}}\right]^{1/4}$$

Radar Cross-Section

The **Radar Cross Section** σ (RCS) is an aircraft-specific quantity that depends on many factors. The computational determination of the RCS is only possible for simple bodies. The RCS of simple geometric bodies depends on the ratio of the structural dimensions of the body to the wavelength.



Practically, the RCS of a target depends on:

- The physical geometry and exterior features of the target,
- The direction of the illuminating radar,
- The radar transmitters frequency,
- The electrical properties of the target's surface.

Whereas in the design of passenger airplanes more attention is paid to effectiveness and safety, in the case of an aircraft used for military purposes, care is taken to ensure that this reflective surface is as small as possible. Measures to achieve this are referred to as *stealth technology*.

What does RCS mean for radar?

The RCS of any reflector can be seen as a ratio to an idealized reference reflector. The projected area of an equivalent isotropic reflector (that is: reflecting equally in all directions) has an RCS of exactly one square meter. Reflecting equally in all directions means: in practice, this is only performed by a spherical reflector with an ideally conducting surface. From a large distance, you cannot see the spherical shape; you can only see a circular area: the so-called projection, both with the same diameter.

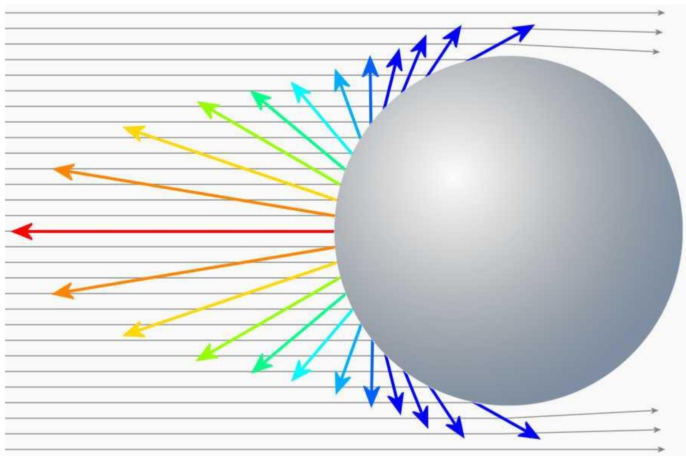


Figure 2: The reference for the radar cross section: a metallic sphere, which in the view offers a (projected) circle with an area of 1 m²

The diameter of this sphere must be approximately 1.128 m to be the size of one square meter. Such an equivalent isotropic reflector delivers the same power per unit measure of solid angle back to the radar, regardless of the aspect angle (i.e., regardless of the direction of the radar). Isotropic reflective does not mean that this sphere would distribute the power arriving from one direction equally in all directions. The shape and size of the distribution are different, but always the same relative to the aspect angle of the illuminating radar, regardless of the angle from which the radar illuminates this sphere. However, the reference reflector can re-radiate only the power it has received:

$$\frac{\boldsymbol{\sigma} \mathbf{S_t}}{\mathbf{4\pi}} = \mathbf{S_r} . \mathbf{r^2} \qquad \text{--- (1)}$$

Where S_t – Power density of the transmitter at the radar target in [W/m²]

S_r – Scattered power density at the receiving site in [W/m²]

σS_t – Power received and re-radiated by the radar target (in watts)

$\frac{\sigma S_t}{4\pi}$ – Power per solid angle, i.e. divided by 4π steradian (in watts/ steradian)

R – Radius of the sphere

This equation expresses only a power balance: only the power that arrives at this reflecting object can be reflected, and this power is radiated in many directions. From this power, however, the radar can only receive a small part again. This part depends on the effective antenna area, the antenna aperture.

Since the power density generated by the radar transmitter and arriving at the reflection point is put into a ratio with the reflected power density arriving at the radar, all other influences, such as free space attenuation and distance to the radar, are eliminated. It is assumed that for a monostatic radar, the propagation conditions are the same on the outbound and return paths.

Let S_r be the power density at the receiving point of the radar, which we now see from the perspective of the reflecting object. This power density has the unit of measurement watts per unit area: W/m². The receiving antenna of the radar has only one effective antenna aperture A_r (this is an area and a part of the surface of a sphere). The received power of the radar antenna is then $S_r \cdot A_r$, which is the power density at the antenna multiplied by its effective aperture.

However, this antenna can receive only a very small portion of the power reflected from the object in all directions, because it occupies only a small part of the surface of the sphere. This area is proportional to the solid angle Ω occupied by the total reflected power distributed on a spherical surface:

$$\boldsymbol{\Omega = \frac{A_r}{r^2}} \qquad \text{--- (2)}$$

Thus, a power density (power per solid angle) of $S_r \cdot A_r / \Omega$ arrives at the receiving antenna. The solid angle can be replaced with the expression from the above equation and leads to

$$\boldsymbol{\frac{S_r A_r}{\Omega} = S_r . r^2} \qquad \text{--- (3)}$$

The expression $S_r \cdot r^2$ thus stands for the received power per unit solid angle (in watts per steradian) and corresponds to equation (1) above. This can then be rearranged to the equation (4) used below.

Non-isotropic reference reflector

In contrast to the isotropic radiator mentioned in the antenna technology, such an isotropic reference reflector can very well be constructed in reality. It would only be very unwieldy because of its dimensions. Since the direction to the radar is known in a measurement setup, a calibrated corner reflector can also be used. As usual in antenna technology, this has a gain G compared to the isotropic reflector, which, however, can be calculated out of the measurement result later.

Computational determination of the reflective area

In the following formula, the radar cross section indicates an effective area that captures the incoming wave and re-radiates it into space. Thus, only a power density caused by the surface of the sphere $(4\pi r^2)$ arrives at the receiving antenna of the radar. The radar cross section σ is defined as:

$$\boldsymbol{\sigma = 4\pi r^2 \frac{S_r}{S_t}} \qquad \text{--- (4)}$$

where σ - apparent area in [m²], measure of the backscattering ability.
S_t - power density of the transmitter at the radar target in [W/m²]
S_r - scattered power density at the receiving location in [W/m²]

The following formulas for calculation of the radar cross section are valid under the condition of optical, i.e., frequency-independent reflection at bodies that are much further away from the radar than the wavelength and which are much larger than the used wavelength of the radar.

Reflection from a sphere	$\boldsymbol{\sigma = \pi r^2}$
Reflection at a cylinder	$\boldsymbol{\sigma_{max} = \frac{2\pi r h^2}{\lambda}}$
Reflection at a flat plate	$\boldsymbol{\sigma_{max} = \frac{4\pi b^2 h^2}{\lambda^2}}$

Reflection at an inclined plate	actually like the previous example, if the projection of the plate to the radar is used as the surface. Only: the reflected energy is reflected in a different direction. So the scanning radar cannot receive this energy. That is why there are bistatic radars, where the emitter and the receivers are spatially separated.

Radar cross sections for point-like targets

In radar technology, point-like targets are targets whose geometric dimensions are smaller than the pulse volume of the radar set. In contrast to volumetric targets (occurring mainly in weather radar), they do not completely fill the pulse volume. In radar signal processing they occupy one resolution cell (maximum two if they are exactly on the boundary).

Targets	**RCS** [m^2]	**RCS** [dB]
bird	0.01	-20
man	1	0
cabin cruiser	10	10
automobile	100	20
truck	200	23
corner reflector	20379	43.1

In reality, the radar cross section is composed of the sum of many small partial powers, which are located at different points of the reflecting object. Depending on the angle from which this object is illuminated, these partial areas have more or less influence, they may be obscured or their distance to the radar may differ by several multiples of half the wavelength so that they overlap partly constructively and partly destructively. The RCS is thus strongly dependent on the aspect angle and can no longer be easily calculated geometrically. It is usually a result of extensive practical measurements either on the original or with a model scaled down to the wavelength currently used in the measurement.

Some targets mentioned as examples have according to their geometrical extension a very large amount of the RCS and reflect therefore a rather large amount of the transmitted energy. The table on the right gives some examples of reflective surfaces in the X-band.

An electronic instrument, which is used for displaying the data visually, is known as display. So, the electronic instrument which displays the information about Radar’s target visually is known as **Radar display**. It shows the echo signal information visually on the screen.

Types of Radar Displays

In this section, we will learn about the different types of Radar Displays. The Radar Displays can be classified into the following types.

A-Scope

It is a two dimensional Radar display. The horizontal and vertical coordinates represent the range and echo amplitude of the target respectively. In A-Scope, the deflection modulation takes place. It is more suitable for **manually tracking Radar**.

B-Scope

It is a two dimensional Radar display. The horizontal and vertical coordinates represent the azimuth angle and the range of the target respectively. In B-Scope, intensity modulation takes place. It is more suitable for **military Radars**.

C-Scope

It is a two-dimensional Radar display. The horizontal and vertical coordinates represent the azimuth angle and elevation angle respectively. In C-Scope, intensity modulation takes place.

D-Scope

If the electron beam is deflected or the intensity-modulated spot appears on the Radar display due to the presence of target, then it is known as blip. C-Scope becomes D-Scope, when the blips extend vertically in order to provide the distance.

E-Scope

It is a two-dimensional Radar display. The horizontal and vertical coordinates represent the distance and elevation angle respectively. In E-Scope, intensity modulation takes place.

F-Scope

If the Radar Antenna is aimed at the target, then F-Scope displays the target as a centralized blip. So, the horizontal and vertical displacements of the blip represent the horizontal and vertical aiming errors respectively.

G-Scope

If the Radar Antenna is aimed at the target, then G-Scope displays the target as laterally centralized blip. The horizontal and vertical displacements of the blip represent the horizontal and vertical aiming errors respectively.

H-Scope

It is the modified version of B-Scope in order to provide the information about elevation angle of the target. It displays the target as two blips, which are closely spaced. This can be approximated to a short bright line and the slope of this line will be proportional to the sine of the elevation angle.

I-Scope

If the Radar Antenna is aimed at the target, then I-Scope displays the target as a **circle**. The radius of this circle will be proportional to the distance of the target. If the Radar Antenna is aimed at the target incorrectly, then I-Scope displays the target as a segment instead of circle. The arc length of that segment will be inversely proportional to the magnitude of pointing error.

J-Scope

It is the modified version of A-Scope. It displays the target as radial deflection from time base.

K-Scope

It is the modified version of A-Scope. If the Radar Antenna is aimed at the target, then K-Scope displays the target as a pair of vertical deflections, which are having equal height. If the Radar Antenna is aimed at the target incorrectly, then there will be pointing error. So, the magnitude and the direction of the pointing error depend on the difference between the two vertical deflections.

L-Scope

If the Radar Antenna is aimed at the target, then L-Scope displays the target as two horizontal blips having equal amplitude. One horizontal blip lies to the right of central vertical time base and the other one lies to the left of central vertical time base.

M-Scope

It is the modified version of A-Scope. An adjustable pedestal signal has to be moved along the baseline till it coincides with the signal deflections, which are coming from the horizontal position of the target. In this way, the target’s distance can be determined.

N-Scope

It is the modified version of K-Scope. An adjustable pedestal signal is used for measuring distance.

O-Scope

It is the modified version of A-Scope. We will get O-Scope, by including an adjustable notch to A-Scope for measuring distance.

P-Scope

It is a Radar display, which uses intensity modulation. It displays the information of echo signal as plan view. Range and azimuth angle are displayed in polar coordinates. Hence, it is called the **Plan Position Indicator** or the **PPI display**.

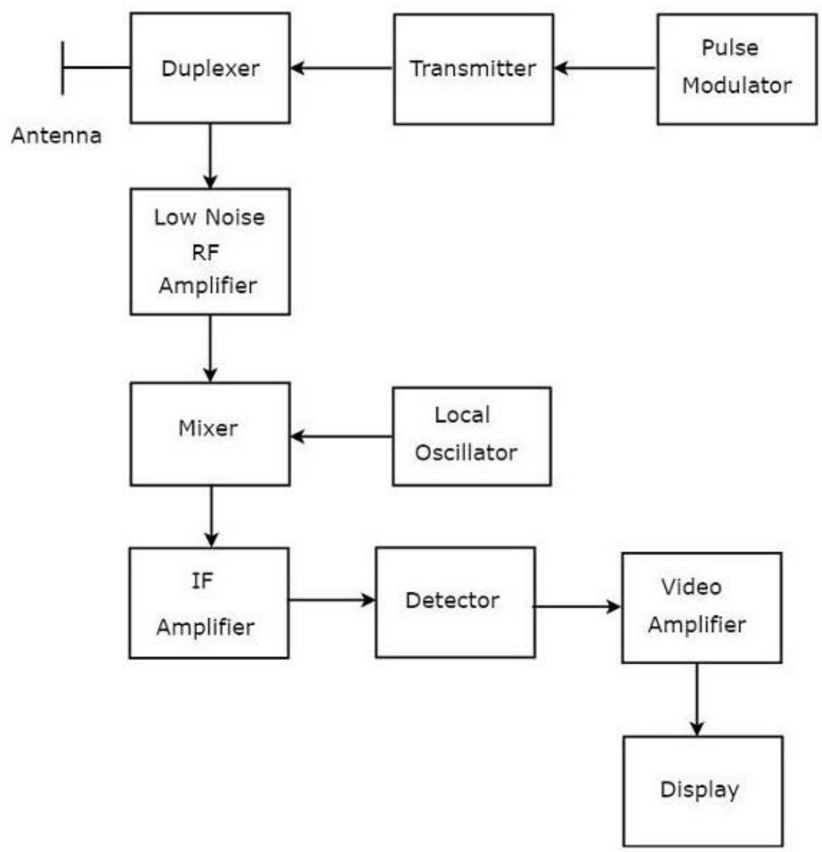
R-Scope

It is a Radar display, which uses intensity modulation. The horizontal and vertical coordinates represent the range and height of the target respectively. Hence, it is called **Range-Height Indicator** or **RHI display**.

Pulse Radar

Block Diagram of Pulse Radar

Pulse Radar uses single Antenna for both transmitting and receiving of signals with the help of Duplexer. Following is the **block diagram** of Pulse Radar.



- **Pulse Modulator** − It produces a pulse-modulated signal and it is applied to the Transmitter.
- **Transmitter** − It transmits the pulse-modulated signal, which is a train of repetitive pulses.

- **Duplexer** − It is a microwave switch, which connects the Antenna to both transmitter section and receiver section alternately. Antenna transmits the pulse-modulated signal, when the duplexer connects the Antenna to the transmitter. Similarly, the signal, which is received by Antenna will be given to Low Noise RF Amplifier, when the duplexer connects the Antenna to Low Noise RF Amplifier.
- **Low Noise RF Amplifier** − It amplifies the weak RF signal, which is received by Antenna. The output of this amplifier is connected to Mixer.
- **Local Oscillator** − It produces a signal having stable frequency. The output of Local Oscillator is connected to Mixer.
- **Mixer** − Mixer can produce both sum and difference of the frequencies that are applied to it. Among which, the difference of the frequencies will be of Intermediate Frequency (IF) type.
- **IF Amplifier** − IF amplifier amplifies the Intermediate Frequency (IF) signal. The IF amplifier shown in the figure allows only the Intermediate Frequency, which is obtained from Mixer and amplifies it. It improves the Signal to Noise Ratio at output.
- **Detector** − It demodulates the signal, which is obtained at the output of the IF Amplifier.
- **Video Amplifier** − As the name suggests, it amplifies the video signal, which is obtained at the output of detector.
- **Display** − In general, it displays the amplified video signal on CRT screen.

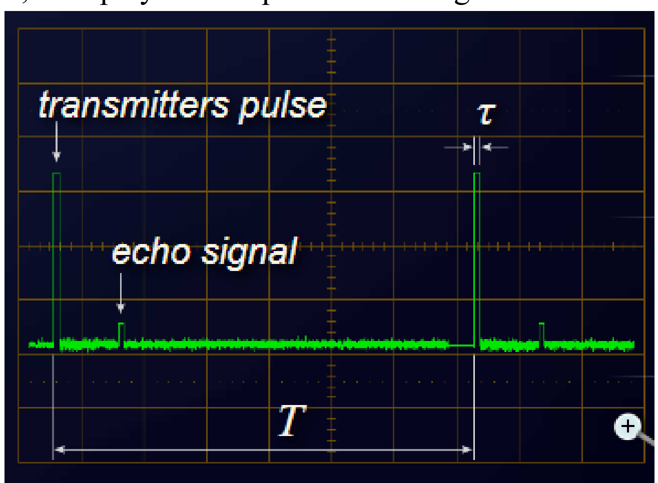


Figure 2: Radar pulse relationships

Pulse radar emits short and powerful pulses and in the silent period receives the echo signals. In contrast to the continuous wave radar, the transmitter is turned off before the measurement is finished. This method is characterized by radar pulse modulation with very short transmission pulses (typically transmit pulse durations of $\tau \approx 0.1 \ \dots \ 1\ \mu s$). Between the transmit pulses are very large pulse pauses $T >> \tau$, which are referred to as the receiving time (typically $T \approx 1\ ms$) as shown in Figure 2. The distance of the reflecting objects is determined by runtime measurement (at a fixed radar) or by comparison of the characteristic changes of the Doppler spectrum with the values for given distances stored in a database (for radar on a fast-moving platform). Pulse radars are mostly designed for long distances and transmit a relatively high pulse power.

Important distinguishing feature to other radar method is the necessary time control of all processes inside the pulse radar. The leading edge of the transmitted pulse is the time reference for the runtime measurement. It ends with the transition of the rising edge of the echo signal in the pulse top. Systematic delays in signal processing must be corrected when calculating the distance. Random deviations influence the accuracy of the pulse radar.

Transmit Signal

The waveform of the transmitted signal can be described mathematically as:

$$\mathbf{s(t) = A(t).\sin[2\pi f(t).t + \varphi(t)]}$$

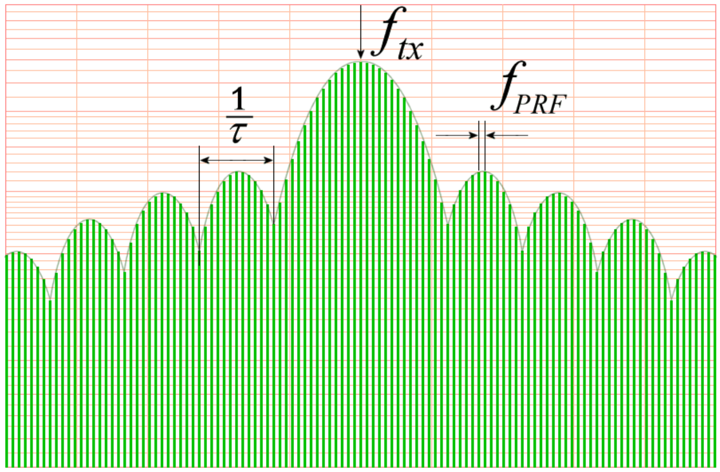


Figure 3: Frequency spectrum of a sequence of rectangular pulses in the vicinity of the transmission frequency f_{tx}

The function *A(t)* is a variation of the amplitude in the function of time *t* - ie an amplitude modulation. In the simplest case, the transmitter is for a short time switched on (for the time τ) and remains in the rest of the time in the "off position". *A(t)* is then in the transmission case = 1, otherwise = 0. The function of time is then determined by the pulse repetition frequency and the duty cycle. Since the radar returns are subject to various losses, an actual amplitude modulation makes little sense except for just this switching function (On/off keying). The envelope of the frequency spectrum of a sequence of rectangular pulses is represented by a $(\sin x)/x$ function. The essential parts of the transmission power (note the logarithmic scale of the ordinate in Figure 3) are in a region $B_{HF} = 2/\tau$ in the vicinity of the transmission frequency f_{tx}.

The pulse repetition frequency f_{PRF} and the duration of the transmitted pulse τ and the receiving time $(T - \tau)$ have an influence on the performance of the radar, e.g. the minimal measuring range (the transmit pulse must have completely exited the antenna) and the maximum unambiguous range (the echo signal must be received in the time before the next transmission pulse). The duration of the transmit pulse τ substantially affects the range resolution ΔR of pulse radar. The range resolution is:

$$\mathbf{\Delta R = 0.5\tau c}$$

The shorter the transmission pulse, the closer one behind the other two reflectors may be positioned so as to be nevertheless detected as two reflectors and not as one large object. The transmitter bandwidth B_{HF} of the pulse radar increases with decreasing pulse width:

$$\mathbf{B_{HF} = \tau^{-1}}$$

The shortening of the pulses limits the maximum range in the case of simple pulse modulation. Under these conditions, the pulse energy E_p can be increased only by the pulse power P_S at a required range resolution.

For the maximum range of the pulse radar, the pulse energy is crucial, and not its pulse power:

$$\mathbf{E_p = P_s\tau = P_{av}T = \frac{P_{av}}{f_{PRF}}}$$

E_p = energy content of the pulse

P_S = transmission pulse power

P_{av} = average power

Significant improvements in this situation can be achieved with the internal modulation of the transmit pulse (intra-pulse modulation). The relationship between the duration of the transmit pulse and the duration of the received pulse is resolved by the pulse compression in the receiver. A location of various reflectors and the measurement of its individual range can also be carried out within the duration of the transmit pulse.

The initial phase of the transmitted signal can either be known and can be predictable (due to the generation of oscillation). In this case, the pulse radar is attributable to the fully coherent radars. The actual phase angle can also be known but the initial state can be unpredictable. Then the radar is one of the pseudo-coherent radars. If this initial phase completely indeterminate (chaotic), then the radar is one of the non-coherent radars. Only with a possible phase-encoded Intra Pulse modulation, this function gets more importance.

Echo Signal

Usually, it is assumed that the duration of the transmitted pulse is equal to the duration of the reflected echo pulse. Thus, in the ratio of the transmitted power and the received power (which is used in the fundamental radar range equation) can be dispensed with a time specification.

- By the reflection of the transmit signal the spectrum may be modified:
- It can occur additional harmonics to the carrier frequency.
- The carrier frequency can be imposed on one or more Doppler frequencies.
- The direction of the polarization can be changed.
- The pulse duration of the echo signal is not constant. The duration of the reflected pulse can be considerably stretched by interference from reflections at areas with slightly different distances (and following different run times).

All together: the echo signal is subject to so many influences that the waveform and the shape of the echo signal in the result must be regarded as unknown. Nevertheless, in order to build an optimal matched receiver or an optimal matched filter, multiple receiving channels must be set up in parallel, taking into account all the possible deformations of the signal. In a selection circuit, the echo signal with the best (*greatest-of*) Signal to Noise Plus Interference Ratio (SNIR) is then further processed. The "position" of the greatest-of-switch is also saved as important information for the identification of this echo signal.

In general, the receiving bandwidth is kept as small as possible, so not much unnecessary noise is received. Therefore, to select the bandwidth only with $B_{HF} = 1/\tau$ for a simple pulse radar. The influence of the noise can be suppressed in the receiver using of pulse integration. Here, a sum of pulse periods is formed. The reflecting object is assumed to be stationary during the time of these pulse periods. Since the noise is randomly distributed, thereby the sum of the noise cannot reach the sum of the echo signals. The signal-to-noise ratio is improved by this measure.

Design, Block Diagram

The construction of pulse radar depends on whether transmitter and receiver are at the same site (monostatic radar) or whether both components are deployed at completely different locations (bistatic radar).

Monostatic pulse radar, in addition to the compact design has the advantage that the important for pulse radars timing devices can be concentrated in a central synchronization block. Internal runtimes of the radar triggers can thus be kept low. An elaborate radar antenna can be used by means of a multiplexer for both transmitting and receiving.

The disadvantage is that often the highly sensitive radar receiver must be switched off by a duplexer for its own protection against the high transmission power. During this time it cannot receive anything.

In a *bistatic pulse radar*, the receiver is equipped with its own antenna in a different location as the transmitter. This has the advantage that the receiver can operate without significant protective measures against to a high transmission power. In the simplest case, a network is constructed from extra receiver locations to existing monostatic pulse radar. The receiving antennas are not very directionally: they must be able to receive from several directions simultaneously. The disadvantage here is the very complex synchronization. Simultaneously with the echo signals, the receiver must also receive the direct transmission signal. From this signal and the known distance to the transmitter, a sync signal must be generated. Principal military application of bistatic configurations are the Over-The-Horizon (OTH) Radars.

The passive radars are a variant of the bistatic radar. They parasitically use a variety of RF emissions (radio or television stations, or external pulse radars) The passive radar calculates the position of the targets from the difference between the time of the direct path of the signal and the additional running time of the reflected echo signals. Ambiguities in the measurement can be excluded on the one hand by direct direction finding involving spurious emissions of the target or by synchronization of two passive radars working at different locations.

Doppler Radar Systems

If the target is not stationary, then there will be a change in the frequency of the signal that is transmitted from the Radar and that is received by the Radar. This effect is known as the **Doppler Effect**.

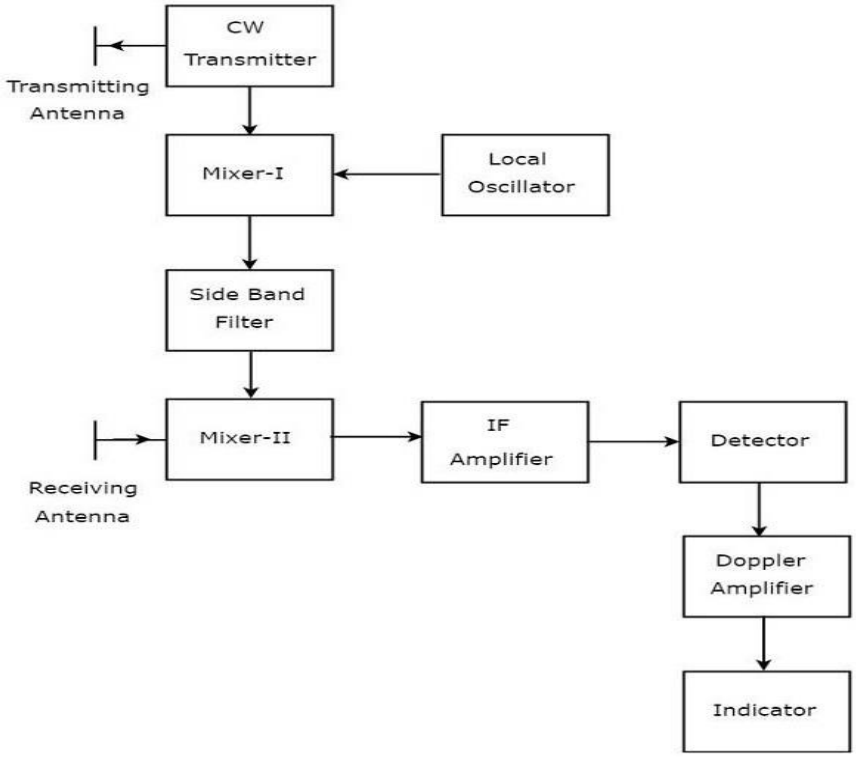
According to the Doppler effect, we will get the following two possible cases −

- The frequency of the received signal will increase, when the target moves towards the direction of the Radar.
- The frequency of the received signal will decrease, when the target moves away from the Radar.

The block diagram of CW Doppler Radar contains a set of blocks and the **function** of each block is mentioned below.

- **CW Transmitter** − It produces an analog signal having a frequency of f_o. The output of CW Transmitter is connected to both transmitting Antenna and Mixer-I.
- **Local Oscillator** − It produces a signal having a frequency of f_l. The output of Local Oscillator is connected to Mixer-I.
- **Mixer-I** − Mixer can produce both sum and difference of the frequencies that are applied to it. The signals having frequencies of f_o and f_l are applied to Mixer-I. So, the Mixer-I will produce the output having frequencies $f_o + f_l$ or $f_o - f_l$.
- **Side Band Filter** − As the name suggests, side band filter allows a particular side band frequencies − either upper side band frequencies or lower side band frequencies. The side band filter shown in the above figure produces only upper side band frequency, i.e $f_o + f_l$.
- **Mixer-II** − Mixer can produce both sum and difference of the frequencies that are applied to it. The signals having frequencies of $f_o + f_l$ and $f_o \pm f_d$ are applied to Mixer-II. So, the Mixer-II will produce the output having frequencies of $f_o + f_l \pm f_d$ or $f_l \pm f_d$.
- **IF Amplifier** − IF amplifier amplifies the Intermediate Frequency (IF) signal. The IF amplifier shown in the figure allows only the Intermediate Frequency, $f_l \pm f_d$ and amplifies it.

- **Detector** − It detects the signal, which is having Doppler frequency, f_d.
- **Doppler Amplifier** − As the name suggests, Doppler amplifier amplifies the signal, which is having Doppler frequency, f_d.
- **Indicator** − It indicates the information related relative velocity and whether the target is inbound or outbound.



Derivation of Doppler Frequency

The distance between Radar and target is nothing but the Range of the target or simply range, R. Therefore, the total distance between the Radar and target in a two-way communication path will be 2R, since Radar transmits a signal to the target and accordingly the target sends an echo signal to the Radar.

If λ is one wave length, then the number of wave lengths N that are present in a two-way communication path between the Radar and target will be equal to $2R/\lambda$.

We know that one wave length λ corresponds to an angular excursion of 2π radians. So, the total angle of excursion made by the electromagnetic wave during the two-way communication path between the Radar and target will be equal to $4\pi R/\lambda$ radians.

Following is the mathematical formula for angular frequency, ω

$$\omega = 2\pi f$$

Following equation shows the mathematical relationship between the angular frequency ω and phase angle ϕ

$$\omega = d\phi\ dt,\ 2\pi f = d\phi\ dt$$

$$f = \frac{1}{2\pi} d\phi\ dt$$

Substitute, $f = f_d$ and $\phi = 4\pi R/\lambda$

$$f_d = \frac{1}{2\pi}\frac{d}{dt}\left(\frac{4\pi R}{\lambda}\right)$$

$$\mathbf{f_d = \frac{1}{2\pi}\frac{4\pi}{\lambda}\frac{dR}{dt}}$$

$$\mathbf{f_d = \frac{2V_r}{\lambda}}$$

Where, f_d is the Doppler frequency

V_r is the relative velocity

We can find the value of Doppler frequency f_d by substituting the values of V_r and λ.

Substitute, λ=C/f

$$\mathbf{f_d = \frac{2fV_r}{C}}$$

Where, f is the frequency of transmitted signal

C is the speed of light and it is equal to 3×10^8m/sec

We can find the value of Doppler frequency, f_d by substituting the values of V_r, f and C

Applications

The applications of radar include the following.

Military Applications

It has 3 major applications in the Military:

- In air defense, it is used for target detection, target recognition, and weapon control (directing the weapon to the tracked targets).
- In a missile system to guide the weapon.
- Identifying enemy locations on the map.

Air Traffic Control

It has 3 major applications in Air Traffic control:

- To control air traffic near airports. The Air Surveillance RADAR is used to detect and display the aircraft's position in the airport terminals.
- To guide the aircraft to land in bad weather using Precision Approach RADAR.
- To scan the airport surface for aircraft and ground vehicle positions

Remote Sensing

It can be used for observing whether or observing planetary positions and monitoring sea ice to ensure a smooth route for ships.

Ground Traffic Control

It can also be used by traffic police to determine the speed of the vehicle, controlling the movement of vehicles by giving warnings about the presence of other vehicles or any other obstacles behind them.

Space

It has 3 major applications

- To guide the space vehicle for a safe landing on the moon
- To observe the planetary systems
- To detect and track satellites
- To monitor the meteors

Summary

- ✶ RADAR stands for Radio Detection and Ranging System. It is basically an electromagnetic system used to detect the location and distance of an object from the point where the RADAR is placed.
- ✶ Pulsed RADAR sends high power and high-frequency pulses towards the target object.

- ✶ The distance between Radar and target is called **Range** of the target or simply range, R.
- ✶ The principle of RADAR detecting moving objects using the Doppler shift works on the fact that echo signals from stationary objects are in the same phase and hence get canceled while echo signals from moving objects will have some changes in phase.
- ✶ It transmits high pulse repetition frequency to avoid Doppler ambiguities. The transmitted signal and the received echo signal are mixed in a detector to get the Doppler shift and the difference signal is filtered using a Doppler filter where the unwanted noise signals are rejected.
- ✶ The **Radar Cross Section** σ (RCS) is an aircraft-specific quantity that depends on many factors. The computational determination of the RCS is only possible for simple bodies. The RCS of simple geometric bodies depends on the ratio of the structural dimensions of the body to the wavelength.
- ✶ The electronic instrument which displays the information about Radar's target visually is known as **Radar display**.
- ✶ If the target is not stationary, then there will be a change in the frequency of the signal that is transmitted from the Radar and that is received by the Radar. This effect is known as the **Doppler Effect**.

Review Questions

Two Marks

1. Define RADAR.
2. What are the basic components of RADAR?
3. What is a RADAR range?
4. Give some types of RADAR display.
5. Define Doppler Effect.
6. What are the uses of RADAR?
7. Give the applications of Doppler RADAR system.
8. Write down the RADAR range equation.
9. What is pulse RADAR?
10. What is bi-static RADAR?
11. Define weather RADAR.
12. Define PPI display
13. What is meant by RHI display?
14. What are the advantages of pulse RADAR system?
15. What are the applications of pulse RADAR system?

Five Marks

1. Explain RADAR.
2. Discuss the types of RADAR.
3. Derive the expression for RADAR range equation.
4. Explain the RADAR cross section.
5. Briefly explain RADAR display.
6. Write notes on Doppler RADAR communication system.

Ten Marks

1. Explain pulsed RADAR communication system.