MSc. Physics Elective_ Instruments and Devices -2, Unit-2

Satellite communication

1.1 Introduction:

A **satellite** is a smaller object that revolves around a larger object in space i. e. moon is a natural satellite of earth. Communication is the exchange of information between two or more entities, through any medium or channel. If the communication takes place between any two earth stations through a satellite, then it is called as **satellite communication**. In this communication, electromagnetic waves are used as carrier signals. These signals carry the information such as voice, audio, video or any other data between ground and space and vice-versa.

1.2 Satellite working mechanism

Satellite communication begins at an earth station that is situated at earth and it is designed to transmit and receive signals from a satellite in an orbit around the earth. Earth stations send the information to satellites in the form of high powered, high frequency (GHz range) signals. The satellites receive the signals and retransmit the signals back to earth where they are received by other earth stations in the coverage area of the satellite. There are some terminology which need to understand for satellite communications as below

Satellite's **footprint** is the area which receives a signal of useful strength from the satellite.

A **repeater** is a circuit, which increases the strength of the received signal and then transmits it. But, this repeater works as a **transponder**. That means, it changes the frequency band of the transmitted signal from the received one.

The frequency with which, the signal is sent into the space is called as **Uplink frequency**. Similarly, the frequency with which, the signal is sent by the transponder to other earth station is called as **Downlink frequency**. This is more clear in the following

The transmission of signal from first earth station to satellite through a channel is called as **uplink**. Similarly, the transmission of signal from satellite to second earth station through a channel is called as **downlink**.

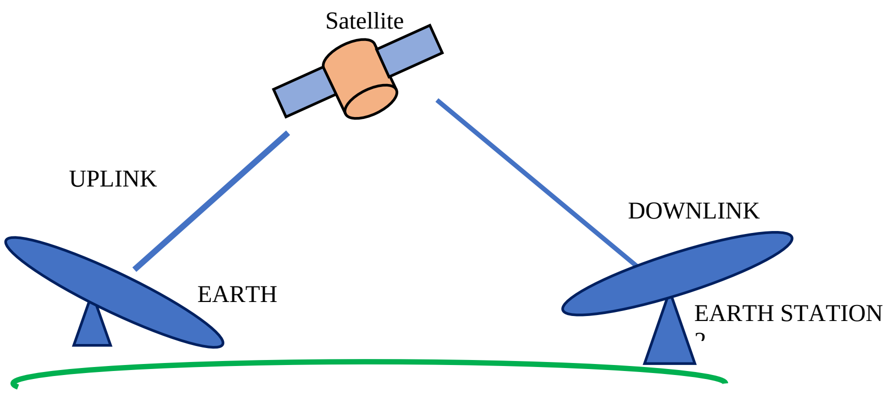


Fig.1. satellite and earth station

1.3 Kepler's Law: Kepler has given three law that suits the satellite communication theory and observations. These are popularly known as **Kepler's laws**.

1.3.1 Kepler's First Law

Kepler's first law states that the path followed by a satellite around the earth will be an **ellipse**. This ellipse has two focal points (foci) F1 and F2 as shown in the figure below. Centre of mass of the earth will always present at one of the two foci of the ellipse.

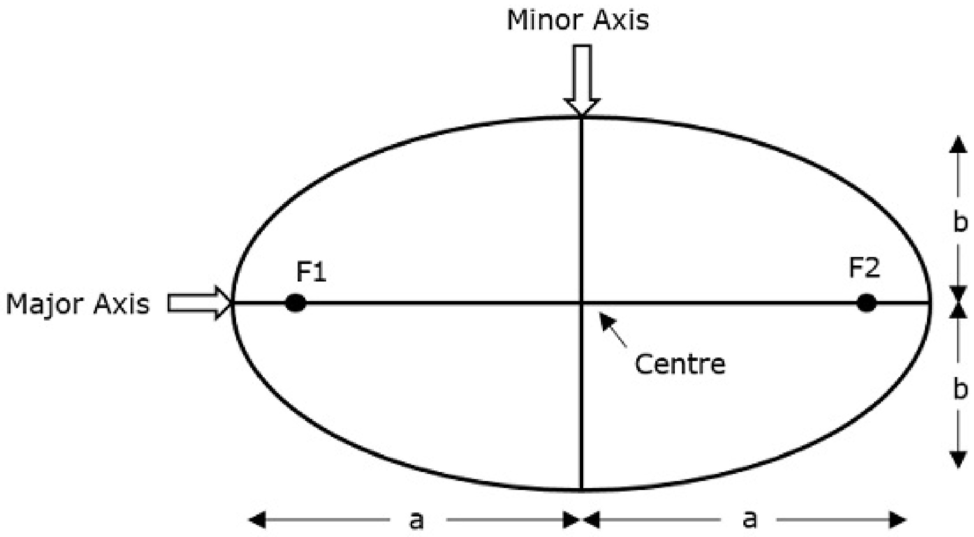


Fig. 2. Kepler's first law

According to Kepler's first Law; If the distance from the centre of the object to a point on its elliptical path is considered, then the farthest point of an ellipse from the centre is called as **apogee** and the shortest point of an ellipse from the centre is called as **perigee**.

Eccentricity "e" of this system can be written as –

$$e = \frac{\sqrt[2]{a^2 - b^2}}{a}$$

Where, **a** & **b** are the lengths of semi major axis and semi minor axis of the ellipse respectively.

For an **elliptical path**, the value of eccentricity (e) is always lie in between 0 and 1, since a is greater than b. Suppose, if the value of eccentricity (e) is zero, then the path will be no more in elliptical shape, rather it will be converted into a circular shape.

1.3.2 Kepler's Second Law

Kepler's second law states that for equal intervals of time, the **area** covered by the satellite will be same with respect to centre of mass of the earth. This is shown in figure below.

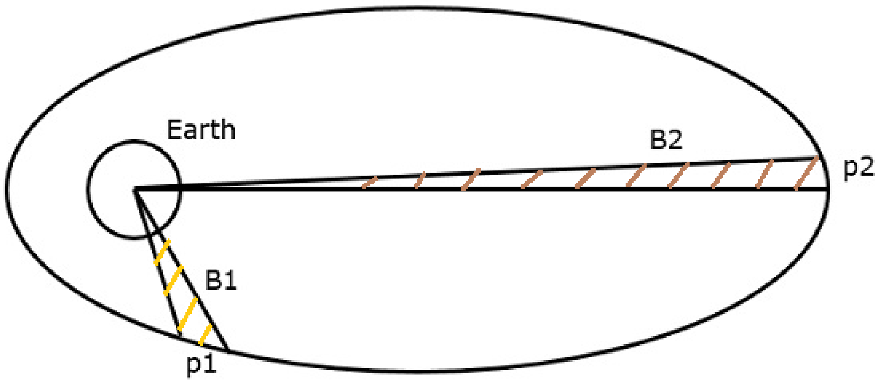


Fig. 3 Kepler's second law

If the satellite covers p1 and p2 distances in the same time interval. Then,

$$B1=B2$$

1.3.3 Kepler's Third Law

Kepler's third law states that, the square of the periodic time of an elliptical orbit is proportional to the cube of its semi major axis length.

$$T^2 \ \alpha \ a^3$$

$$T^2 = \frac{4\pi^2}{\mu} (a)^3$$

Where $\frac{4\pi^2}{\mu}$ is the proportionality constant,

μ is Kepler's constant and its value is equal to 3.986005 x 10^{14} m^3 /sec^2

n is the mean motion of the satellite in radians per second. That can be written as

$$n= \frac{2\pi}{T}$$

Note − A satellite, when it revolves around the earth, undergoes a pulling force from the earth, which is gravitational force. Similarly, it experiences another pulling force from the sun and the moon. Therefore, a satellite has to balance these two forces to keep itself in its orbit.

1.4 Geostationary Orbit:

A Geostationary Orbit (GO) is a circular orbit, located directly above the earth's equator at an altitude of about 35,786 kilometers. This orbit has zero eccentricity and zero inclination. There is only one geostationary orbit and it should have a period equal to the earth rotation period. A satellite rotate in the GO is rotating in the direction of earth's (west to east) rotation and with same angular velocity as the earth's rotation. So that, this satellite is seems stationary when seen from the surface of the earth.

Every Geostationary orbit is a Geo-synchronous orbit. But, the converse need not be true. Weather monitoring satellites are geostationary satellites as they need to have a constant view of the same area. Telecommunication and television satellites are some of the satellites in the geostationary orbit, as you want them to continuously transmit to the antenna on the Earth's surface, without changing the direction of the antenna.

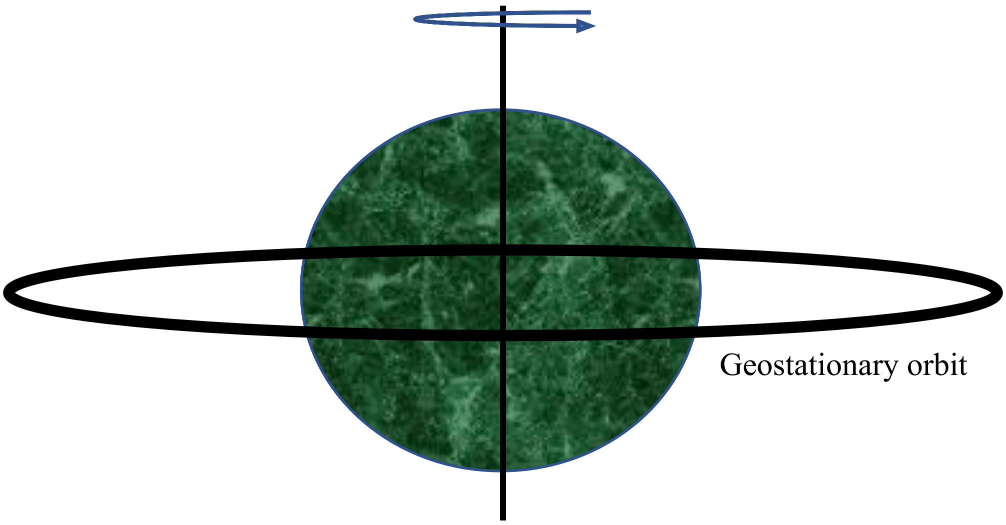


Fig. 4 Geostationary satellite

Kepler' third law are used to find the radius of the orbit (for a circular orbit, the semimajor axis is equal to the radius).

a_{GSO}= 42164 km

The equatorial radius of the earth, to the nearest kilometer, is

a_E = 6378 km

hence the height is

h= a_{GSO}-a_E = 42164-6378= 35786 Km

This value is often rounded up to 36,000 km for approximate calculations.

1.5 Power System:

Any system to operate in regular manner some power is needed. In the same way huge internal power is needed for continuous operation of satellites. These system can be solar cell, solar array solar panel and rechargeable batteries etc. These are continues sources of power as they are naturally gaining energy.

1.5.1 Solar Cells

Basically, the **solar cells** produce electrical power (current) from incident sunlight. Therefore, solar cells are used primarily in order to provide power to other subsystems of satellite. The radiation on a satellite from the sun has an intensity averaging about 1.4 kW/m^2. A single solar cell is not useful as it generates very less power so solar arrays are needed to provide electrical power to satellites. Solar cells are used to generate power to run the sensor, active heating and cooling and telemetry. They generate power to spacecraft propulsion electric propulsion. There are two types of arrays are used

1.5.2 Cylindrical solar arrays are used in spinning satellites. Only part of the cylindrical array will be covered under sunshine at any given time. Due to this, electric power gets generated from the partial solar array. This is the drawback of this type.

The drawback of cylindrical solar arrays is overcome with **Solar sail**. This one produce more power because all solar cells of solar sail are exposed to sun light.

1.5.3 Rechargeable batteries

Rechargeable batteries are needed at the time of eclipses because during this time system do not get the power from sun light. So, in that situation the other subsystems get the power from **rechargeable batteries**. These batteries produce power to other subsystems during launching of satellite also and other types of sun blockage. In general, these batteries charge due to excess current, which is generated by solar cells in the presence of sun light.

1.6 Attitude control

The attitude of a satellite refers to its orientation in space with respect to earth. Attitude control is necessary so that the antennas, are pointed correctly towards Earth. Attitude control is necessary, for example, to ensure that directional antennas point in the proper directions. In the case of earth environmental satellites, the earth-sensing instruments must cover the required regions of the earth, which also requires attitude control. There are several forces can interact to affect the attitude of the spacecraft; 1. Gravitational forces from the sun, moon, and planets; 2.Solar pressures acting on the spacecraft body, antennas or solar panels; 3. earth’s magnetic field. So Altitude and Orbit Control **(AOC)** subsystem are capable of placing the satellite into the right orbit, whenever it is deviated from the respective orbit. We can understand it in two parts

1. Attitude control subsystem
2. Orbit control system

Attitude control system: Altitude control subsystem takes care of the orientation of satellite in its respective orbit. It can be understand by three axis system.

a) Roll axis
b) Yaw axis
c) Pitch axis

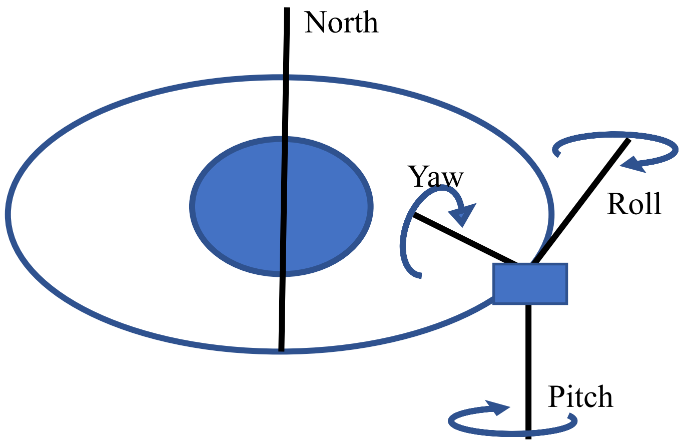


Fig. 5 Attitude control System

Roll axis is in the direction in which satellite is moving in the orbital plane, yaw axis is in the direction to centre of earth and pitch is the direction perpendicular to the orbital plane. All three axes pass through the centre of gravity of the satellite. For an equatorial orbit, movement of the satellite about the roll axis moves the antenna footprint north and south; movement about the pitch axis moves the footprint east and west; and movement about the yaw axis rotates the antenna footprint.

Orbit control subsystem

Orbit control subsystem is useful in order to bring the satellite into its correct orbit, whenever the satellite gets deviated from its orbit. If there is any change in satellite orbit, then it sends a signal regarding the correction to orbit control subsystem. Then, it will resolve that issue by bringing the satellite into the correct orbit. In this way, the AOC subsystem takes care of the satellite position in the right orbit and at right altitude during entire life span of the satellite in space.

1.7: Multiple access method

Sometimes a satellite's service is present at a particular location on the earth station and sometimes it is not present. That means, a satellite may have different service stations of its own located at different places on the earth. They send carrier signal for the satellite. In this situation, we do multiple access to enable satellite to take or give signals from different stations at time without any interference between them. Following are the **three types** of multiple access techniques.

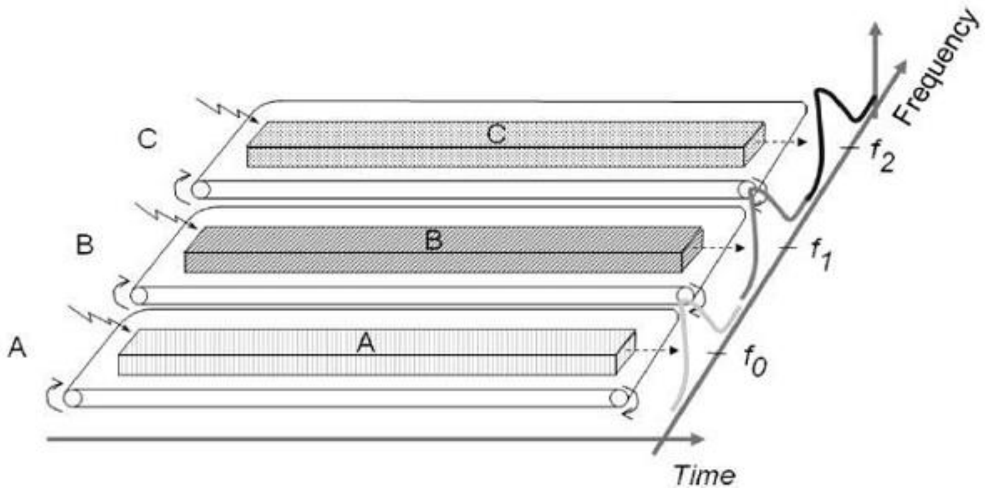
- FDMA (Frequency Division Multiple Access)
- TDMA (Time Division Multiple Access)
- CDMA (Code Division Multiple Access)

Now, let us discuss each technique one by one.

1.7.1: Frequency Division Multiple Access (FDMA)

In this type of multiple access, we assign each signal a different type of frequency band (range). So, any two signals should not have same type of frequency range. Hence, there won't be any interference between them, even if we send those signals in one channel.

One perfect **example** of this type of access is our radio channels. We can see that each station has been given a different frequency band in order to operate.



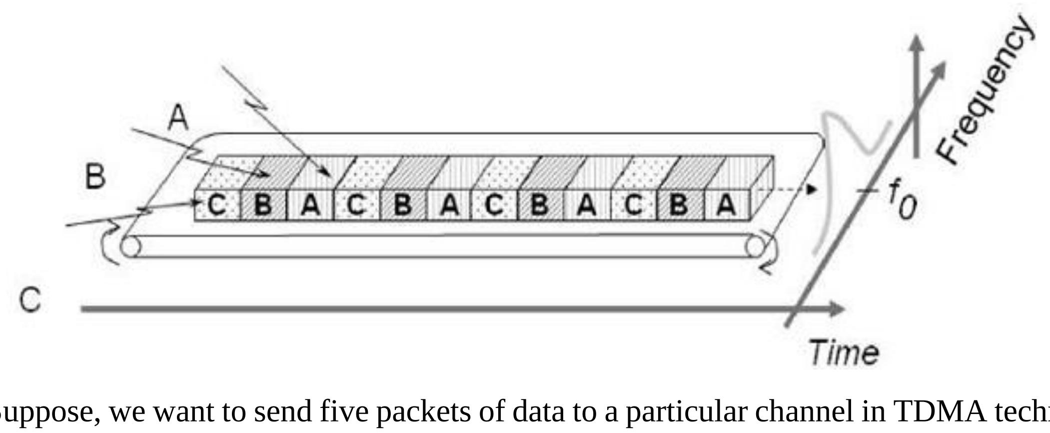
Let's take three stations A, B and C. We want to access them through FDMA technique. So we assigned them different frequency bands.

As shown in the figure, satellite station A has been kept under the frequency range of 0 to 20 HZ. Similarly, stations B and C have been assigned the frequency range of 30-60 Hz and 70-90 Hz respectively. There is no interference between them.

The main **disadvantage** of this type of system is that it is very burst. This type of multiple access is not recommended for the channels, which are of dynamic and uneven. Because, it will make their data as inflexible and inefficient.

1.7.2: Time Division Multiple Access (TDMA)

In time division multiple access, we give certain time frame to each channel. Within that time frame, the channel can access the entire spectrum bandwidth. Each station got a fixed length or slot. The slots, which are unused will remain in idle stage.



Suppose, we want to send five packets of data to a particular channel in TDMA technique. So, we should assign them certain time slots or **time frame** within which it can access the entire bandwidth. It can be understand like few packets are active, which transmits data. Whereas, few packets are idle because of their non-participation. This format gets repeated every time we assign bandwidth to that particular channel.

Although, we have assigned certain time slots to a particular channel but it can also be changed depending upon the load bearing capacity. That means, if a channel is transmitting heavier loads, then it can be assigned a bigger time slot than the channel which is transmitting lighter loads. This is the biggest **advantage** of TDMA over FDMA. Another advantage of TDMA is that the power consumption will be very low.

Note − In some applications, we use the **combination** of both **TDMA and FDMA** techniques. In this case, each channel will be operated in a particular frequency band for a particular time frame. In this case, the frequency selection is more robust and it has greater capacity over time compression.

1.7.3: Code Division Multiple Access (CDMA)

In CDMA technique, each signal is associated with a particular code that is used to spread the signal in frequency and/or time. All such signals will be received simultaneously at an earth station, but by using the key to the code, the station can recover the desired signal by means of correlation. The other signals occupying the transponder channel appear very much like random noise to the correlation decoder

Our cellular system works on the CDMA technique. We can see that no two persons' mobile number match with each other although they are same X or Y mobile service providing company's customers using the same bandwidth.

The basic **advantage** of this type of multiple access is that it allows all users to coexist and use the entire bandwidth at the same time. Since each user has different code, there won't be any interference.

In this technique, a number of stations can have number of channels unlike FDMA and TDMA. The best part of this technique is that each station can use the entire spectrum at all time.

Both FDMA and TDMA can be operated as pre assigned or demand assigned systems. CDMA is a random-access system, there being no control over the timing of the access or of the frequency slots accessed

1.8: Antenna look angle

Earth station will receive the maximum signal level, if it is located directly under the satellite. Otherwise, it won't receive maximum signal level and that signal level decreases as the difference between the latitude and longitude of earth station increases.

So, based on the requirement we can place the satellite in a particular orbit. Now, let us discuss about the look angles.

1.8.1: Look Angles

The following two angles of earth station antenna combined together are called as **look angles**.

- Azimuth Angle
- Elevation Angle

Generally, the values of these angles change for non-geostationary orbits. Whereas, the values of these angles don't change for geostationary orbits. Because, the satellites present in geostationary orbits appear stationary with respect to earth.

These two angles are helpful in order to point at the satellite directly from the earth station antenna. So, the **maximum gain** of the earth station antenna can be directed at satellite. We can **calculate** the look angles of geostationary orbit by using longitude & latitude of earth station and position of satellite orbit.

Azimuth Angle

The angle between local horizontal plane and the plane passing through earth station, satellite and centre of earth is called as **azimuth angle**.

The **formula** for Azimuth angle (α) is

$$\alpha=180^{o}+\tan^{-1}(\tan G/\tan L)$$

Where,

- ***L*** is Latitude of earth station antenna.
- ***G*** is the difference between position of satellite orbit and earth station antenna.

The following **figure** illustrates the azimuth angle.

Measure the **horizontal angle** at earth station antenna to north pole as shown in figure below. It represents azimuth angle. It is used to track the satellite horizontally.

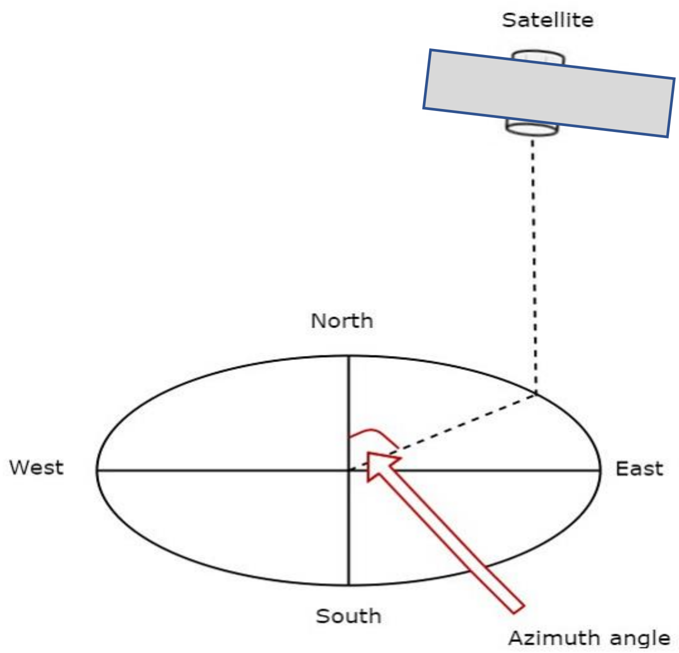


Fig. 7 Azimuthal angle

Elevation Angle

The angle between vertical plane and line pointing to satellite is known as Elevation angle. Vertical plane is nothing but the plane, which is perpendicular to horizontal plane. Measure the **vertical angle** at earth station antenna from ground to satellite as shown in the figure. It represents elevation angle

The Elevation angle β is

$$\beta = tan^{-1}(\frac{cosG.cosL - 0.15}{\sqrt{1 - cos^2G.cos^2L}})$$

We can calculate the elevation angle by using above formula. The following **figure** illustrates the elevation angle.

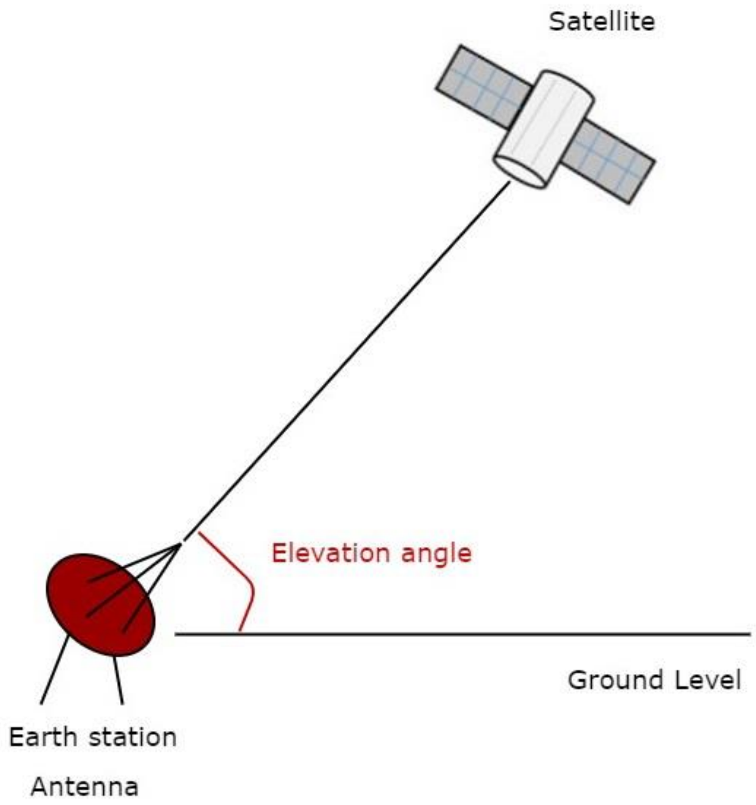


Fig. 8 Elevation Angle

1.9: Limit of visibility

There will be east and west limits on the geostationary arc visible from any given earth station. The limits will be set by the geographic coordinates of the earth station and the antenna elevation. The lowest elevation in theory is zero, when the antenna is pointing along the horizontal. A quick estimate of the longitudinal limits can be made by considering an earth station at the equator, with the antenna pointing either west or east along the horizontal, as shown in Fig. 9. The limiting angle is given by

$$\Theta = \cos\text{-}1(a_{GSO}/a_E)$$

$$=\cos^{-1}(6378/42164) = 81.3$$

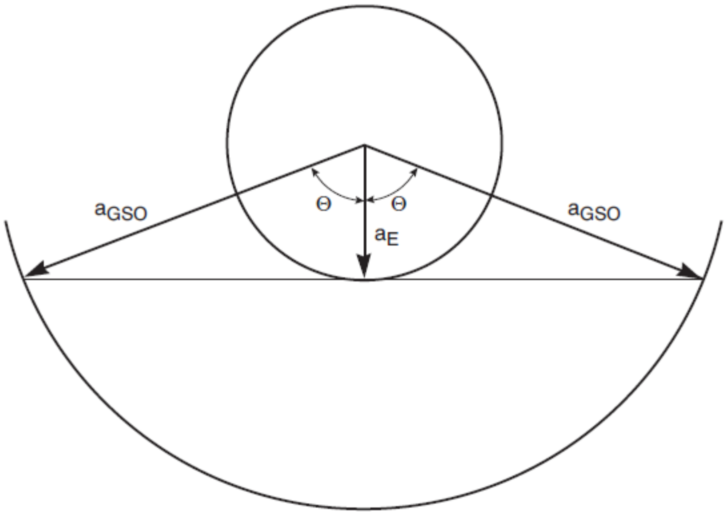


Fig. 9 Limits of Visibility

Thus, for this situation, an earth station could see satellites over a geostationary arc bounded by 81.3° about the earth-station longitude. In practice, to avoid reception of excessive noise from the earth, some finite minimum value of elevation is used, which will be denoted here by E_{lmin}. A typical value is 5°. The limits of visibility will also depend on the earth-station latitude. Let S represent the angle subtended at the satellite when the angle

$$\sigma_{min} = 90° + E_{lmin}.$$

Applying the sine rule gives

$$S = \sin^{-1}\{(R/a_{GSO})\sin \sigma_{min}\}$$

A sufficiently accurate estimate is obtained by assuming a spherical earth of mean radius 6371 km, angle S is known, angle b is found from

$$b = 180^{o} - \sigma_{min} - S$$

$$B = \cos^{-1}(\cos b/\cos\lambda_E)$$

Once angle B is found, the satellite longitude can be determined from

$$B = \phi_E - \phi_S$$

Earth station latitude λ_E
Earth Station longitude ϕ_E
Longitude of Subsatellite point

Example 3.4 Determine the limits of visibility for an earth station situated at mean sea level, at latitude 48.42° north, and longitude 89.26 degrees west. Assume a minimum angle of elevation of 5°.

Solution Given data:
$\lambda_E = 48.42°$; $\phi_E = -89.26°$; $El_{\min} = 5°$; $a_{GSO} = 42164$ km; $R = 6371$ km

$$\sigma_{\min} = 90° + El_{\min}$$

Equation (3.17) gives:

$$S = \arcsin\left(\frac{6371}{42164}\sin 95°\right) = 8.66°$$

Equation (3.18) gives:

$$b = 180 - 95° - 8.66° = 76.34°$$

Equation (3.19) gives:

$$B = \arccos\left(\frac{\cos 76.34°}{\cos 48.42°}\right) = 69.15°$$

The satellite limit east of the earth station is at

$$\phi_E + B = \underline{\underline{-20°}} \text{ approx.}$$

and west of the earth station at

$$\phi_E - B = \underline{\underline{-158°}} \text{ approx.}$$

1.10: Transponder

The subsystem, which provides the connecting link between transmitting and receiving antennas of a satellite is known as **Transponder**. It is one of the most important subsystem of space segment. The transponder is the series of components that provides the communications channel, or link, between the uplink signal received at the uplink antenna, and the downlink signal transmitted by the downlink antenna. The key elements of the payload portion of the space segment, is the transponder and antenna subsystems.

Transponder performs the functions of both transmitter and receiver (Responder) in a satellite. Hence, the word 'Transponder' is obtained by the combining few letters of two words, Transmitter **(Trans)** and Responder **(ponder)**.

Transponder performs mainly **two functions**. Those are amplifying the received input signal and translates the frequency of it. In general, different frequency values are chosen for both uplink and down link in order to avoid the interference between the transmitted and received signals.

The **block diagram** of transponder is shown in below figure.

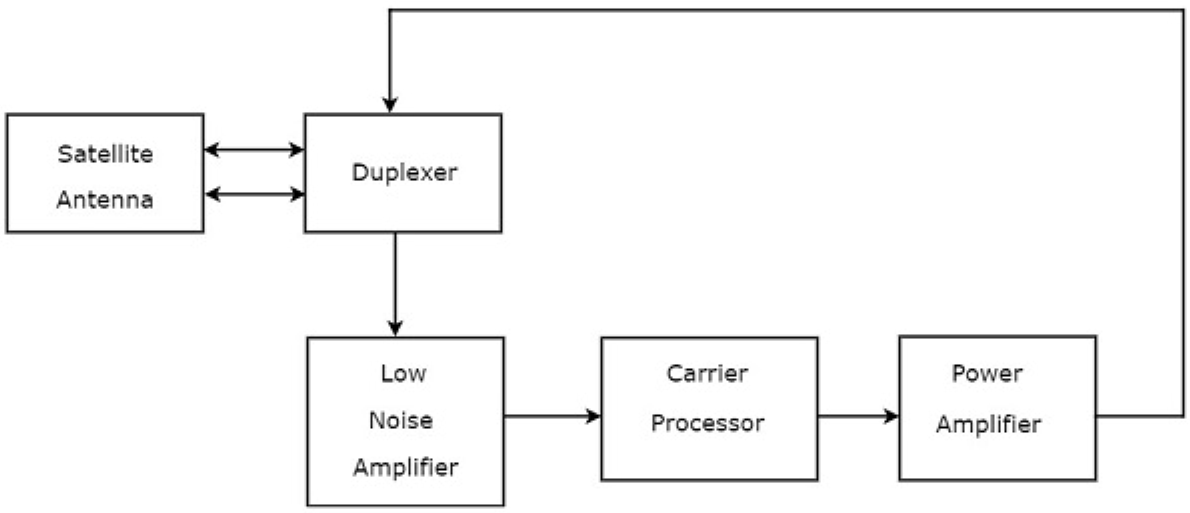


Fig. 10. Transponder block diagram

We can easily understand the operation of Transponder from the block diagram itself. The function of each block is mentioned below.

- **Duplexer** is a two-way microwave gate. It receives uplink signal from the satellite antenna and transmits downlink signal to the satellite antenna.
- **Low Noise Amplifier** (LNA) amplifies the weak received signal.
- **Carrier Processor** performs the frequency down conversion of received signal (uplink). This block determines the type of transponder.
- Power Amplifier amplifies the power of frequency down converted signal (down link) to the required level.

1.10.1:Types of Transponders

Basically, there are **two types** of transponders. Those are Bent pipe transponders and Regenerative transponders.

Bent Pipe/Conventional Transponders

Bent pipe transponder receives microwave frequency signal. It converts the frequency of input signal to RF frequency and then amplifies it. It is suitable for both analog and digital signals.

Regenerative/Processing Transponders

Regenerative transponder performs the functions of Bent pipe transponder. i.e., frequency translation and amplification. In addition to these two functions, Regenerative transponder also performs the demodulation of RF carrier to baseband, regeneration of signals and modulation. It is suitable only for digital signals. The main **advantages** of Regenerative transponders are improvement in Signal to Noise Ratio (SNR) and have more flexibility in implementation.

1.11: Satellite wide band receiver

The wideband receiver is shown in figure below. A duplicate receiver is provided so that if one fails, the other is automatically switched in. The combination is referred to as a redundant receiver, meaning that although two are provided, only one is in use at a given time. The first

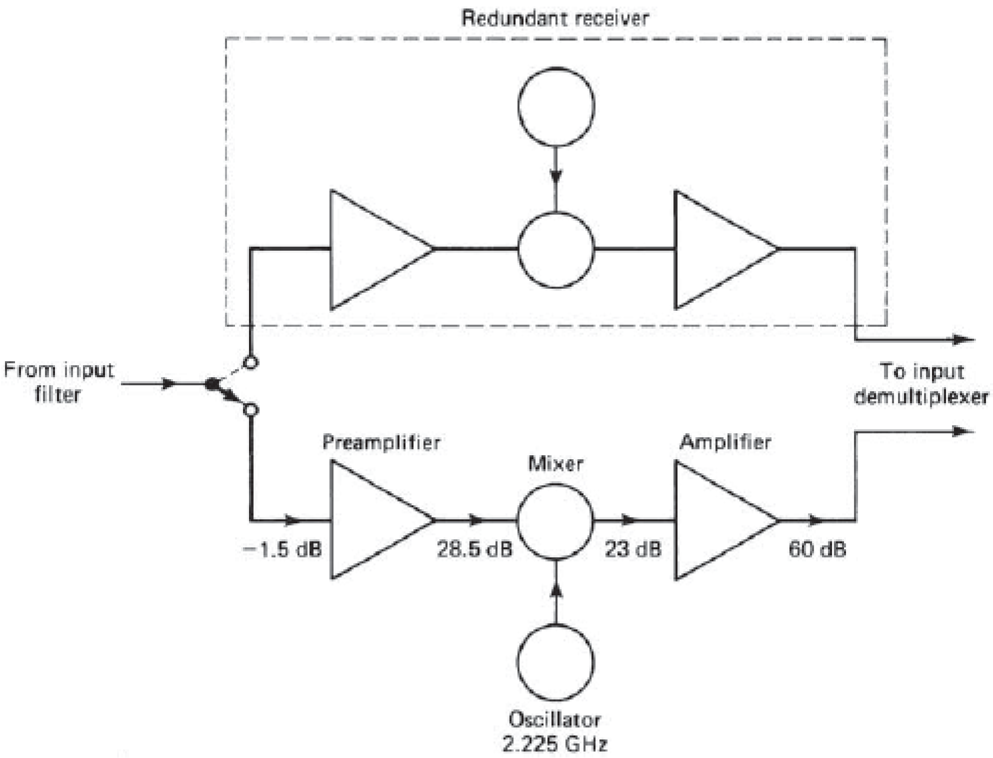


Fig.11 Wide band receiver

stage in the receiver is a low-noise amplifier (LNA). This amplifier adds little noise to the carrier being amplified, and at the same time it provides sufficient amplification for the carrier to override the higher noise level present in the following mixer stage.

The LNA feeds into a mixer stage, which also requires a local oscillator (LO) signal for the frequency-conversion process. The power drive from the LO to the mixer input is about 10 dBm. The oscillator frequency must be highly stable and have low-phase noise. A second amplifier follows the mixer stage to provide an overall receiver gain of about 60 dB. The Splitting the gain between the preamplifier at 6 GHz and the second amplifier at 4 GHz prevents oscillation, which might occur if all the gain were to be provided at the same frequency. The wideband receiver utilizes only solid-state active devices. In some designs, tunnel-diode amplifiers have been used for the preamplifier at 6 GHz in 6/4-GHz transponders and for the parametric amplifiers at 14 GHz in 14/12-GHz transponders. With advances in field-effect transistor (FET) technology, FET amplifiers, which offer equal or better performance, are now available for both bands.

2. Optical Sources

2.1 Incandenscent Sources

The incandescent light bulb or lamp is a source of electric light that works by incandescence, which is the emission of light caused by heating the filament. They are made in an extremely wide range of sizes, wattages, and voltages. Incandescent lamps are often considered the least energy efficient type of electric lighting commonly found in residential buildings. Although inefficient, incandescent lamps possess a number of key advantages--they are inexpensive to buy, turn on instantly, are available in a huge array of sizes and shapes and provide a pleasant, warm light with excellent color rendition.

There are three common types of incandescent lamps used in residential applications:

- Standard incandescent or pear-shaped A-19 lamps

  These are "A" type lamp are standard incandescent bulbs are the least efficient light source commonly found in homes. These lamps produce visible light by heating a tiny coil or filament of tungsten wire that glows when it is heated by an electrical current

- Energy-saving or halogen A-19 lamps

  A halogen lamp is a type of incandescent lamp with a capsule that holds a special halogen gas composition around the heated filament to increase the efficacy of the incandescence lamp. They are more energy efficient than standard incandescent bulbs but somewhat more costly. Halogen lamps may also have a special inner coating that reflects heat back into the capsule to further improve efficacy by "recycling" the otherwise wasted heat. Together, the filling and coating recycle heat to keep the filament hot with less electricity. They also provide excellent color rendition. They are effectively less costly in term of consumption of light.

- Reflector or parabolic reflector (PAR) lamps, sometimes called "flood" or "spot" lamps Reflector bulbs spread the direct light over specific areas. They are used mainly for floodlighting, spotlighting, and down lighting applications both indoor and outdoor.

2.2 Luminescence:

Luminescence is any emission of light (electromagnetic waves) from a substance that does not arise from heating like incandescent sources. The word luminescence is derived from the Latin word for light, *lumen,* and the Latin, *escentia,* meaning 'the process of' and hence is the process of giving off light. The simple meaning of luminescence is the giving off light. The best example is the moon which is giving light at night by taking light from sun. luminescence is the emission of electromagnetic radiation from the material/surface after the excitation. Luminescence may be seen in neon and fluorescent lamps; television, radar, and X-ray fluoroscope screens; organic substances such as luminol or the luciferins in fireflies and glowworms etc. In all these phenomena, light emission does not result from the material being above room temperature, and so luminescence is often called cold light. The practical value of luminescent materials lies in their capacity to transform invisible forms of energy into visible light. Based upon the means of **excitation** there are various types of luminescence are exist

(a) Photoluminescence; produced by absorption of light or photons,

(b) Radioluminescence produced by the impact of higher energy particles from cosmic rays or radioactive matter,

(c) Electroluminescence produced by the application of an electric field,

(d) Mechano or triboluminescence produced by the application of mechanical force,

(e) Cathodoluminescence produced by cathode rays or electrons,

(f) Chemiluminescence produced by chemical process,

(g) Sonoluminescence produced by high frequency sound waves or phonons

(h) Bioluminescence produced by biological processes

Photoluminescence:

Photoluminescence is the process of **emission** of light by the **absorption** of photon is called the photoluminescence.

(a) Absorption: it is the process of absorbing the incident energy from photon. Let us consider that the energy level of ground state electrons or lower energy state (occupied) electrons is E_1 and the next higher energy level or higher energy state is E_2. The electrons jumps from ground state (E_1) to the excited state or higher energy level (E_2) after absorbing some energy from photon. The electrons in the higher energy level are called excited electrons.

(b) Excitation: It is a process where a photon of energy E gives up some or all of its energy to raise electrons from occupied low energy levels E_1 to unoccupied higher energy levels E_2. Interaction of ionizing radiation with the solid can transfer sufficient energy to electrons in the valence band for transferring them to the conduction band. A good number of these electrons return to the ground state by radiative (causing fluorescence or phosphorescence) or non-radiative (heat) light emission.

c). **Emission: (**Radiative transitions) A photon can be emitted when an electron drops from an upper (higher) energy level to lower energy level. These levels can be intrinsic band states or impurity levels. Radiative transitions are of different category here we define fluorescence and phosphorescence.

Fluorescence and Phosphorescence: They are radiative emission transitions from the higher to the lower energy states. The classification of the two is based on the persistence of emission after the source of exciting energy is removed. Fluorescence and phosphorescence are having wide applications in various field. Many substances continue to luminesce for extended periods even after the exciting energy is cut off. The delayed emission is generally called phosphorescence and emission during the time of excitation is called fluorescence. More clearly the emission with in a lifetime of 10^{-8} to 10^{-5} is indisputably fluorescence and the long-lived emissions are phosphorescence (life time of the excited state being of the order of 10^{-8} for atomic dipole emission and is of the order 10^{-5} s for phosphorescent sulphides like ZnS) The various processes that are important in luminescence analysis are shown in the Fig. 2.1. If a molecule does not undergo a photochemical reaction after it absorbs electromagnetic radiation in the UV or visible region, it is normally promoted to a vibrational level in the excited electronic singlet state. The time for a molecule to pass from the ground state to an electronically excited state is about 10^{-15} S. Since this time period is so short, the atomic nuclei in the molecule do not appreciably change their relative positions. So as per Franck Condon Principle, electronic transitions essentially occur without change in positions of nuclei. Immediately after excitation by electromagnetic radiation, a molecule has the same geometry and is in the same environment as it was in the ground state. The upper vibrational levels for an excited electronic state of the molecule relax vibrationally in about 10^{-12} S. With in this period the excited molecule relaxes vibrationally to the lowest vibrational level of the lowest excited singlet state.

The vibrational relaxation process is accompanied by a loss in thermal energy. Once the molecule is in the lowest vibrational level of the lowest excited singlet state, a radiative transition can occur in which the molecule drops to one of a number of possible vibrational levels of the ground electronic state. This radiative transition is called fluorescence. These almost always occur from the lowest vibrational level of the lowest excited singlet state in a molecule. The decay time of fluorescence is about the same order of magnitude as the lifetime of an excited singlet state namely 10^{-9} to 10^{-7} S.
As shown in the figure, other processes can compete with fluorescence. The excited molecule can loose energy by other means such as internal conversion and inter-system crossing. Internal conversion is a radiation less process. However, a radiationless transition between the first excited singlet state to the lowest excited triplet state is called inter-system crossing. A change in spin occurs with inter-system crossing, and thus spin selection rules cannot be obeyed rigorously. If inter system can favourably compete with fluorescence or internal conversion to ground state, the molecule can pass from lowest excited singlet state to a triplet state. Then the molecule will undergo vibrational relaxation to arrive at the lowest vibrational level of the lowest excited triplet state. From this state, molecule can undergo radiative transition or undergo inter-system crossing to the ground state. This radiative transition results in phosphorescence. Since phosphorescence originates from the lowest triplet state in a molecule, its decay time is similar to the lifetime of the triplet state, which is about $l0^{-4}$ to10 seconds. The rate of phosphorescence is relatively slow compared to other processes associated with excited molecules.

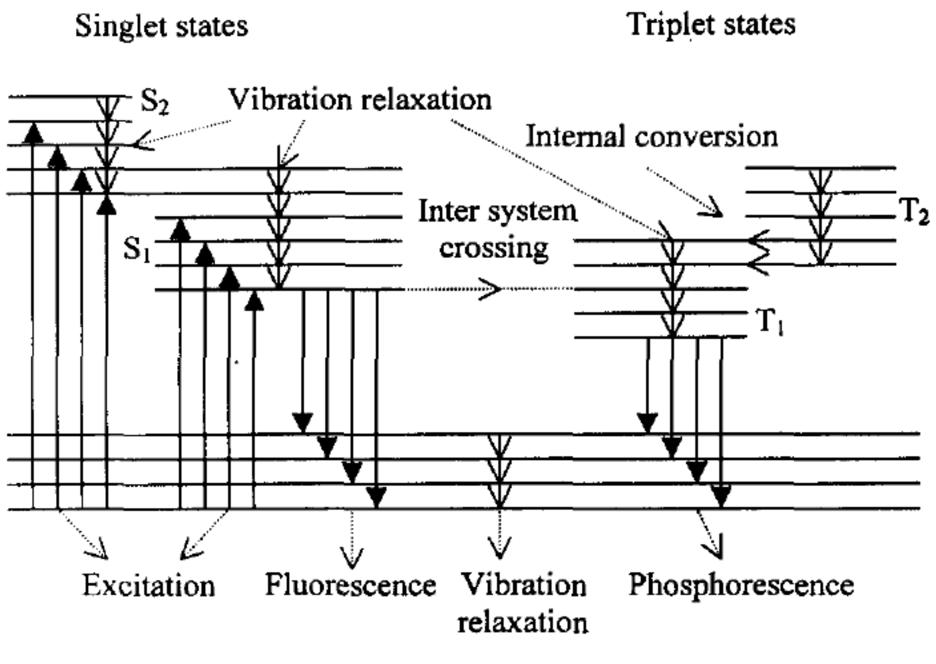


Fig. 2.1 Partial energy level diagram for a photoluminescent molecule

Fluorescence is prompt photoluminescence that occurs very shortly after photoexcitation of a substance, while phosphorescence is long-lived photoluminescence that continues long after the photoexcitation has ceased.

Chathodoluminescence

Cathodoluminescence (CL) is the emission of photons of characteristic wavelengths from a material that is under high-energy electron bombardment. The electron beam is typically produced in an electron microprobe or scanning electron microscope or in a cathodoluminesce microscopy attachment to a petrographic microscope.

If a crystal is bombarded by electrons with sufficient energy, electrons from the lower-energy valence band are promoted to the higher-energy conduction band. When the energetic electrons attempt to return to the ground state valence band, they may be temporarily trapped (on the scale of microseconds) by intrinsic (structural defects) and/or extrinsic (impurities) traps. If the energy lost when the electrons vacate the traps is emitted is in the appropriate energy/wavelength range, luminescence will result. Most of the photons fall in the visible portion of the electromagnetic spectrum (wavelengths of 400-700 nm) with some falling in the ultraviolet (UV) and infrared (IR) portions of the electromagnetic spectrum

Electroluminescence, production of light by the flow of electrons, as within certain crystals. Electroluminescence is one of the few instances in which a direct conversion of electric energy into visible light takes place without the generation of heat like the incandescent lamp.

There are two distinct mechanisms that can produce electroluminescence in crystals:
pure or intrinsic and charge injection. The principal differences between the two mechanisms are that in the first, no net current passes through the phosphor (electroluminescent material) and in the second, luminescence prevails during the passage of an electric current.

In **intrinsic electroluminescence**, thermal activation and the electric field liberate atomic electrons into the conduction band. Many of these conduction electrons are accelerated by the field until they collide with luminescent centres, and ionizing them (i.e., ejecting electrons from their atoms). Light is emitted in the normal way as soon as an electron recombines with an ionized atom of the centre. Because the effect dies away when constant voltage is applied, an alternating voltage may be used to create a sustained light emission.

Electroluminescence can also result from **charge injection**, as when an electrode contacts a crystal to provide a flow of electrons or holes or a voltage is applied to a *p*–*n* junction causing a current to flow; i.e., electrons flow from the *n*-type material into the *p*-type material. In both cases, the electrons lose energy when recombining with centres or holes accompanied by the emission of light.

LED

2.3 LED Structures and Working

Light Emitting Diode (LED) works only in forward bias condition. In forward biased, the free electrons from n-side and the holes from p-side are pushed towards the junction. When free electrons reach the junction or depletion region, some of the free electrons recombine with the holes. Because of the recombination of free electrons and holes in the depletion region, the width of depletion region decreases. As a result, more charge carriers will cross the p-n junction. Thus, recombination takes place in depletion region as well as in p-type and n-type semiconductor. The free electrons in the conduction band releases energy in the form of light before they recombine with holes in the valence band. Light is produced by the solid state process that is called electroluminescence. The emitted light is depends on the band gap of semiconductor material.

LED structure: A typical LED structure is shown in the figure below. The active region is the region that emit radiation on the recombination of electrons and holes on the application of forward bias energy. A junction is formed between p-type and n-type semiconductor. A substrate is needed to design LED. If semiconductor of same material is used then it form homojunction LED structure. That is low efficient LED because of its structure. If we can confine the electrons and holes to a small region (like a potential well) then it is possible to increase radiative recombination efficiency. For this heterojunction LED is proposed in which different band gap semiconductor material are used. The refractive index of any material depends on the band gap of the material. Higher band gap material has lower refractive index. It means LED from heterojunction can be formed by engineering the dielectric wave guide in the device and loss of emitted energy can be tailored. Hence efficiency is increased in heterostructure LED.

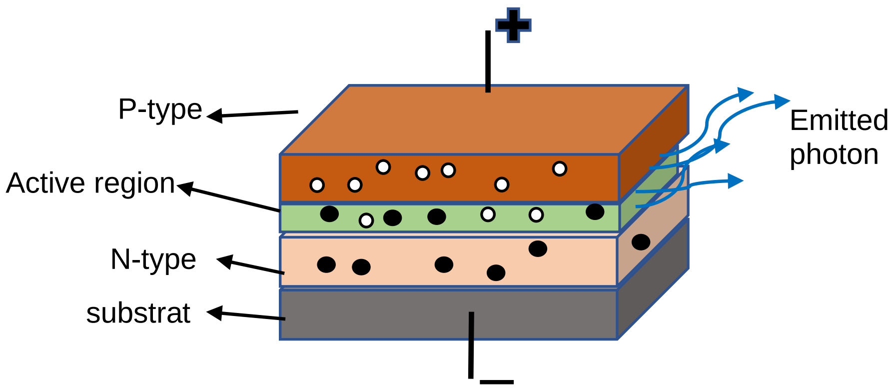


Fig. 2.2 Schematic of LED

Drawback of homojunction LED:

A p-region must be narrow to allow to photons to escape without much reabsorption. The consequence of this is some of the injected electron in the p-side reach the surface by diffusion and recombine through crystal defects near the surface this radiation less recombination process decrease the light output. If the recombination occur over a large distance due to large electron diffusion length then the reabsorption of emitted photon become higher. Hence the efficiency will be decreased.

Colors of an LED

The material used for constructing LED determines its color. It can be say that the wavelength or color of the emitted light depends on the band gap the material.

- Gallium Arsenide (GaAs) – infra-red
- Gallium Arsenide Phosphide (GaAsP) – red to infra-red, orange
- Aluminium Gallium Arsenide Phosphide (AlGaAsP) – high-brightness red, orange-red, orange, and yellow
- Gallium Phosphide (GaP) – red, yellow and green
- Aluminium Gallium Phosphide (AlGaP) – green
- Gallium Nitride (GaN) – green, emerald green
- Gallium Indium Nitride (GaInN) – near ultraviolet, bluish-green and blue
- Silicon Carbide (Sic) – blue as a substrate
- Zinc Selenide (ZnSe) – blue
- Aluminium Gallium Nitride (AlGaN) – ultraviolet

2.4 LED Parameters:

Radiance: It is defines as the measure of optical power radiated into a unit solid angle per unit area of emitting surface. High radiance is necessary to couple sufficiently high optical power level into a fiber or other device.

Quantum efficiency:

Internal quantum efficiency:

$$\eta_o = \frac{number\ of\ photon\ generated\ internally}{number\ of\ carrier\ passing\ through\ the\ junction}$$

$$\eta_o = \frac{R_r}{R_r + R_{nr}}$$

$$\eta_o = \frac{1/\tau_r}{1/\tau_r + 1/\tau_{nr}}$$

Where If the radiative recombination rate per unit volume is R_r and nonradiative recombination rate is R_{nr}. They are inversely related to their associated lifetimes. The internal efficiency η_o is the ratio of radiative recombination rate to the total recombination rate.

***External Quantum efficiency*:**

$$\eta_{ext} = \frac{number\ of\ photon\ emitted\ externallly}{number\ of\ carrier\ passing\ the\ junction}$$

$$\eta_{ext} = \eta_o \eta_{op}$$

where η_{op} is the optical efficiency of the system. This is related to the device optics and how the light is extracted out of the device.

Power Efficiency:

$$\eta_p = \frac{optical\,power\,out}{input\,power}$$

$$\eta_p = \frac{number\ of\ photon\ \times\ hv}{I\ V}\ \approx\ \eta_{ext}$$

Transient Response Time:

Emission response time is the time delay between the application of current pulse and the onset of optical emission. This time is the factor is limiting the bandwidth with which the source can be modulated directed by the injected current.

Modulation capability:

Modulation of output of LED is necessary to code the input as pulses this is done by switching the current on and off. If the drive current of the LED is modulated at frequency ώ then the optical output power of the device will vary as

$$P(\acute{\omega}) = \frac{Po}{\sqrt{1 + \acute{\omega}\tau_i}}$$

Where τ_i is the recombination lifetime and P_o is the power emitted at zero modulated frequency

Maximum possible modulated frequency depends largely on the recombination lifetime of the electron hole pair in the recombination region of the device. This time can vary with different LED types.

Modulation bandwidth can be defined by the optical and electrical terms as shown in figure 2.3 below. Modulation bandwidth is defined where the electrical signal power is dropped to half of its maximum (Po) power at zero modulation frequency as the frequency increases, this is the electrical 3-dB point. Sometimes optical modulation is defined in terms of optical output power P(ώ) where the optical power goes to half of the input optical power at zero modulation as the frequency increases. This point is called the optical 3-dB point. It is defined in figure below in terms of current ratio of the output and input.

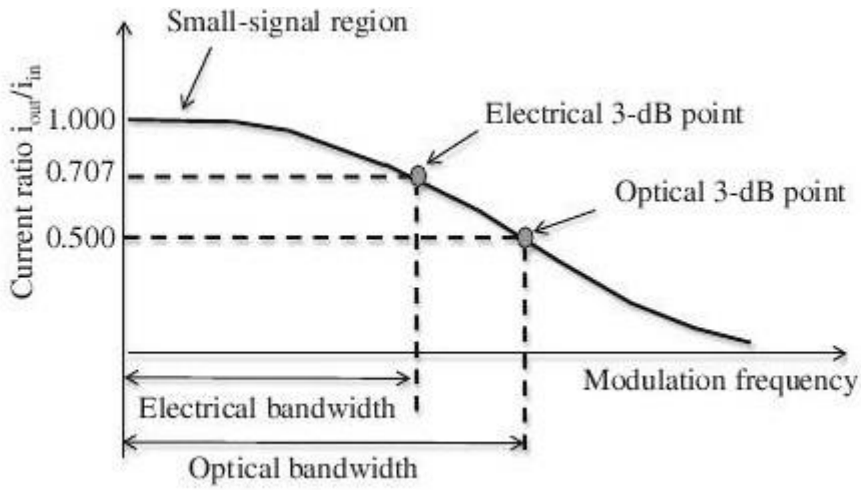
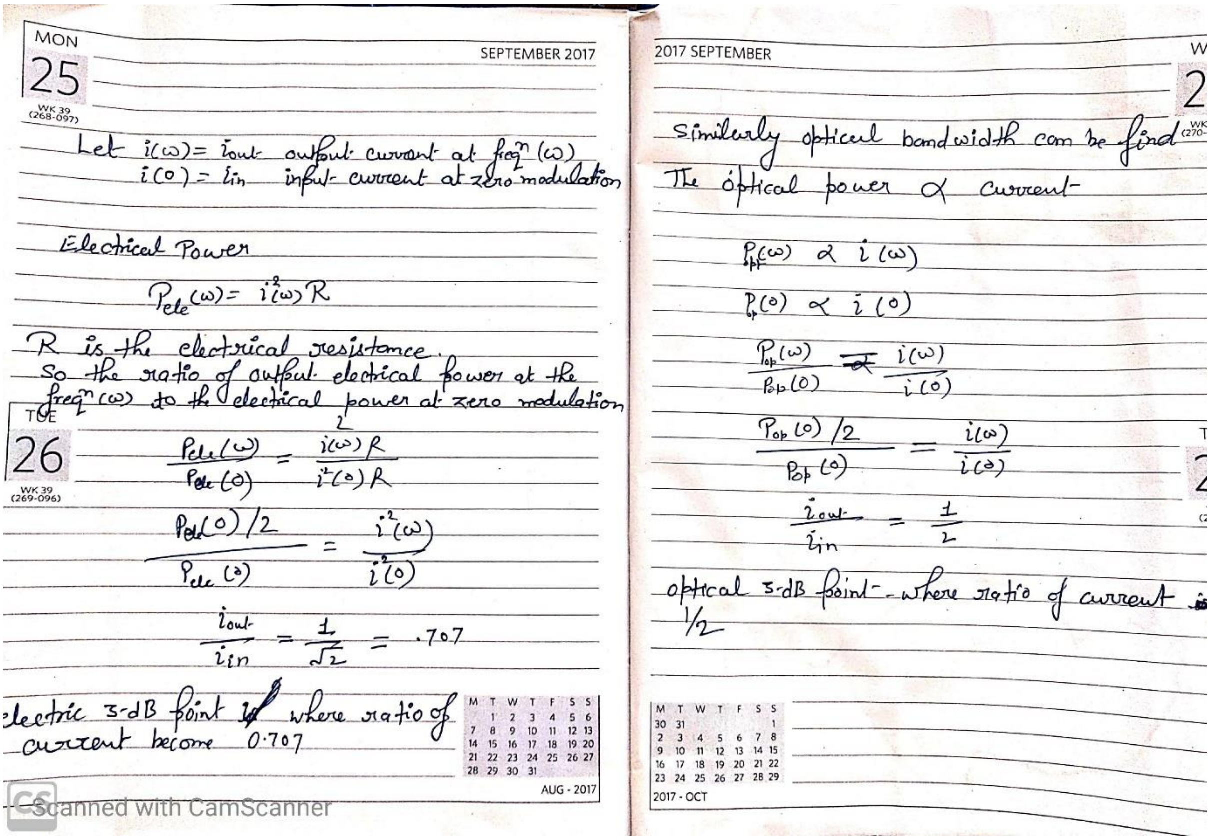


Fig. 2.3 electrical and optical bandwidth



MON 25 SEPTEMBER 2017 WK 39 (268-097)

Let $i(\omega) = i_{out}$ output current at freq^n (ω)

$i(0) = i_{in}$ input current at zero modulation

Electrical Power

$$P_{ele}(\omega) = i^2(\omega) R$$

R is the electrical resistance.

So the ratio of output electrical power at the freq^n (ω) to the electrical power at zero modulation

TUE 26 WK 39 (269-096)

$$\frac{P_{ele}(\omega)}{P_{ele}(0)} = \frac{i^2(\omega) R}{i^2(0) R}$$

$$\frac{P_{ele}(0)/2}{P_{ele}(0)} = \frac{i^2(\omega)}{i^2(0)}$$

$$\frac{i_{out}}{i_{in}} = \frac{1}{\sqrt{2}} = .707$$

electric 3-dB point where ratio of current become 0.707

AUG - 2017

2017 SEPTEMBER

Similarly optical bandwidth can be find

The optical power $\propto$ current

$$P_{op}(\omega) \propto i(\omega)$$

$$P_{op}(0) \propto i(0)$$

$$\frac{P_{op}(\omega)}{P_{op}(0)} = \frac{i(\omega)}{i(0)}$$

$$\frac{P_{op}(0)/2}{P_{op}(0)} = \frac{i(\omega)}{i(0)}$$

$$\frac{i_{out}}{i_{in}} = \frac{1}{2}$$

optical 3-dB point - where ratio of current is 1/2

2017 - OCT

2.5 LED Characteristics:

Two important characteristics of a LED are its Light intensity with Current, and Junction Voltage with Current.

1) **Light Intensity (Optical Power) vs. Current**

This is a very important characteristic of an LED. It can be defined as the optical power generated by an LED is directly proportional to the injected current *I* (current through the LED). However, in practice the characteristic is generally non-linear, especially at higher currents. The near-linear light output characteristic of an LED is exploited in small length fiber optic analog communication links, such as fiber optic closed-circuit TV.

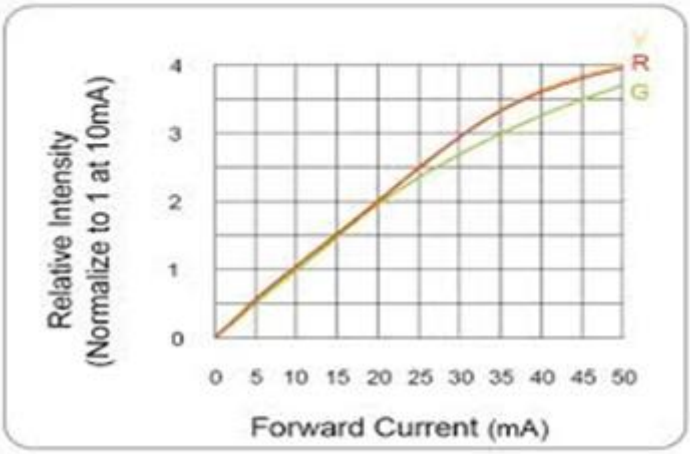


Fig 2.4

2) Junction Voltage with current:

The junction voltage vs. current characteristic of an LED is similar to the V-I characteristics of diodes. However, there is one major difference. The knee voltage of a diode is related to the barrier potential of the material used in the device. Silicon diodes and bipolar junction transistors are very commonly used whose knee voltage or junction voltage is about 0.7 V. Very often it is wrongly assumed that other diodes also have the same junction voltage. In an LED, depending on the material used its junction voltage can be anywhere between 1.5 to 2.2 Volts.

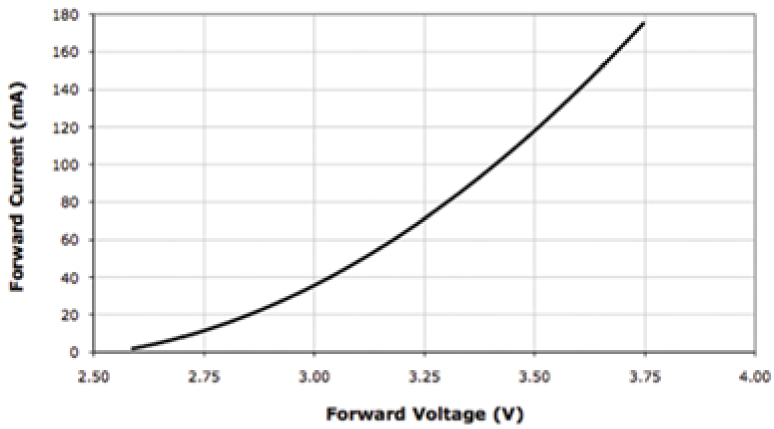


Fig. 2.5

2.6 Applications of Light Emitting Diodes

- LED is used as a bulb in the homes and industries
- The light emitting diodes are used in the motorcycles and cars, traffic signals etc.
- These are used in the mobile phones to display the message.
- They are used in digital computers, digital watches, microprocessors, multimeters etc.
- LED are used in Automotive heat lamps, Camera flashes, Aviation lighting .

Advantages of LED’s

- The cost of LED’s is less and they are tiny and light in weight.
- By using the LED’s the electricity is controlled.
- The intensity of the LED differs with the help of the microcontroller.
- The brightness of light emitted by LED is depends on the current flowing through the LED. Hence, the brightness of LED can be easily controlled by varying the current. This makes possible to operate LED displays under different ambient lighting conditions.
- LED’s have longer lifetime and can emit different colors of light.
- LED’s operates very fast. They can be turned on and off in very less time.
- LED’s do not contain toxic material like mercury which is used in fluorescent lamps.

Disadvantages of LED

1. LED’s need more power to operate than normal p-n junction diodes.
2. Luminous efficiency of LED is low.

2.6 Semiconductor Laser:

Semiconductor lasers work in a similar manner, with the requirement of an optical cavity for feedback, stimulated emission and laser gain. Consider a simple p-n junction manufactured from a material doped to have excess holes and a material doped to have excess electrons. When a junction is formed with these two materials in equilibrium, a voltage develops that prevents electrons in the conduction band of the n-type material from diffusing across the barrier and combining with holes in the p-type material. When a voltage equal to this potential is applied across the device, current flows and electrons combine with holes, producing photons in the process. In most laser diodes, degenerately doped semiconductor materials are used. Degenerately doped means that the Fermi levels (the statistical point where 50% of the electrons will be found) are actually within the valence (for p-type material) and conduction (for n-type material) bands themselves. Application of a voltage across the gap causes the Fermi levels for each type of material, aligned at equilibrium, to split into two distinct levels separated by the applied voltage. Electrons in the conduction band of the n-type material now lie just below the Fermi level of that material, and holes in the valence band of the p-type material lie just above the Fermi level of that material. An inversion is hence generated since there are more electrons in the upper energy band than in the lower band. Of course, electrons and holes involved in the recombination process can lie anywhere in these bands, so a range of wavelengths are possible, with the longest wavelengths corresponding to the bandgap energy.

Electrons are confined to bands, the top band being defined by the Fermi level for the n-type material and the bottom of the conduction band Ec, and the lower band being defined by the Fermi level for the p-type material and the top of the valence band Ev. Photons with energies corresponding to jumps within these bands encounter amplification by stimulated emission since a population inversion exists, but when the energy of an incident photon exceeds the energy corresponding to the difference between the Fermi levels, it is absorbed rapidly by electrons in the valence band of the p-type material to promote these to the conduction band. The material is thus strongly absorbing at wavelengths shorter than the energy corresponding to the difference between the Fermi levels, as illustrated in Figure 2.6. Optical gain by stimulated emission can, therefore, occur only for photons with a specific range of energies. Shorter wavelengths are absorbed and longer wavelengths simply lack the energy to make the transition.

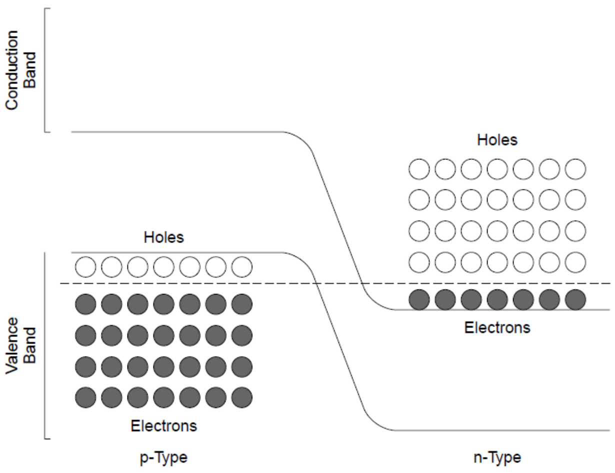


Fig. 2.6 Energy level in semiconductor

2.6.1 Laser Structure:

Homojunction laser

The simplest structure for a laser diode is the homojunction laser diode, which uses a single junction. These are fabricated of a single junction between two direct-bandgap materials of the same type, one p-type and one n-type, that is called a homojunction since both materials are of the same type. Light is emitted by electron–hole pair recombinations in the thin active region formed by the junction of the two materials (the depletion region). Usually, gallium arsenide (GaAs) is used, with each part of the device doped slightly differently: one part with an electron donor and one part with an electron acceptor. Mirrors for the laser cavity are fabricated simply by cleaving the crystal at right angles to the laser axis. Having an index of refraction of 3.7, the reflectivity of each mirror may be calculated to be 33% by using the Fresnel equations. This represents a large loss in the cavity; however, most semiconductor laser materials have ample gain, to allow such a simple configuration.

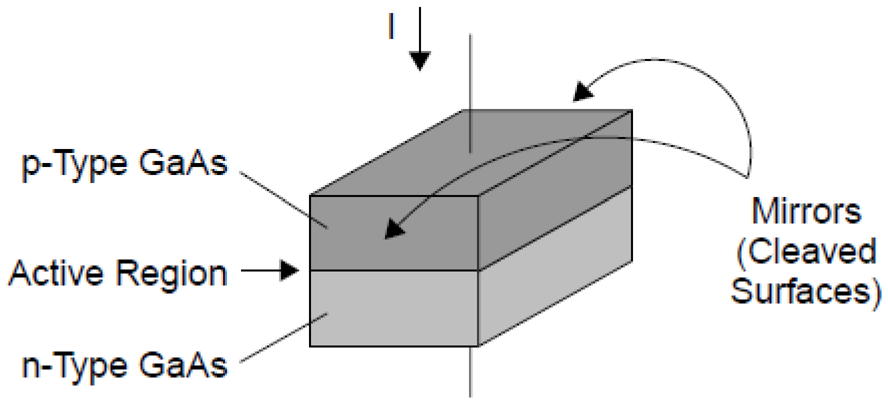


Fig. 2.7 Homojunction laser structure

Heterojunction Laser:
Improvements in the structure of the laser may be made by confining the Intracavity beam in a dielectric waveguide structure formed from the semiconductor material itself. Such a structure requires two interfaces of different indexes of refraction, one on top and one below the active region, so two junctions are formed in what is called a heterostructure laser diode, or in this

case, a double heterostructure, since there are two confining interfaces. To obtain different indexes of refraction, two different materials are required. GaAs is generally used as the higher-index material and aluminum–gallium–arsenide (AlGaAs) as the lower-index material. As depicted in Figure 2.8, AlGaAs is doped to form p- and n-type materials which essentially have identical indexes of refraction given that dopant concentration is small. Between layers of these materials, GaAs is sandwiched as the active-region material, from which laser light is emitted. Differences between the indexes of refraction occurring at each interface form a reflector that confines light inside the GaAs layer, which drastically improves efficiency, and more important, lowers threshold current for the device by increasing gain. The active region (GaAs) is typically only 0.1 *m*m in thickness. Usually, a stripe contact is used on the top of the structure to make an electrical connection to the device. This further limits the area of the active region in the GaAs laser (since current is not spread out over the top surface area of the entire structure), which serves to increase current density and further lower threshold current. In a real laser diode of this type,
more than three layers are generally required, and a layer that serves as an electrical interface between metal contacts and each AlGaAs layer is usually employed.

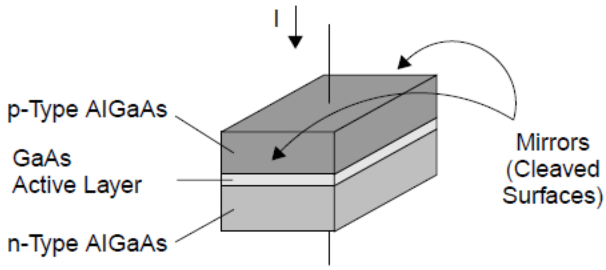


Fig.2.8 Heterostructure of Laser

The double-heterostructure arrangement confines intracavity light in only one direction (top and bottom) of the GaAs layer, but a further improvement in performance can be made by manufacturing the device so that a confining dielectric interface exists on all four sides of the active region in a buried heterostructure

2.6.2 Characteristics of Semiconductor Laser:

The output beam from a laser diode consists of an elliptically shaped beam which is diffraction limited. Thinner dimensions, such as the thickness of the active layer, have higher divergence since gain in that direction comes primarily from photons which are confined by the top and bottom layers of the structure and have been reflected many times. The wider active layer of the device means fewer reflections and hence lower divergence in that direction. Usually, a stripe contact on the top layer restricts the flow of current through the device, increasing current density in the center of the active region.

The wavelength stability of a laser diode depends highly on the temperature of the diode. As the temperature of a laser diode increases, the refractive index of the semiconductor material itself changes. Since the resonant wavelength of a cavity depends on the refractive index of the material, the wavelength shifts toward longer wavelengths as the temperature of the device increases. The output wavelength of a conventional laser diode increases more or less linearly as temperature does.

The output of a conventional laser diode consists of many closely spaced longitudinal modes which lead collectively to a large spectral bandwidth in the output of the device. Even with a single-mode laser, temperature can affect the output spectra, since as temperature changes, the wavelength at which the cavity is resonant shifts and the laser can hop between allowed longitudinal modes. Again, temperature control is required to ensure wavelength stability.

Types of Semiconductor Laser

There are several type of semiconductor lasers in which some name are given below:

1. Fabry Perot (FP) Lasers
2. Distributed Feedback (DFB) lasers
3. Multi Quantum Well (MQW) laser
4. External Cavity Diode Lasers (ECDL)
5. Vertical Cavity Surface Emitting Lasers (VCSEL)

Vertical cavity surface emitting laser (the VCSEL). In a laser of this type, light is not emitted from the edge of the device but rather through the entire top layer of the semiconductor crystal itself. While edge emitting laser diodes produce an elliptically shaped output beam that has high divergence, requiring an external lens to collimate it into a usable beam, a VCSEL produces a round beam of much higher quality. Rather than emission from the edge of the diode, light is emitted from the surface of a VCSEL. In addition to a better beam shape, VCSELs feature single longitudinal mode operation with a narrow spectral linewidth. A VCSEL features a thin active layer (100 to 200 nm) like that of a conventional laser diode, but whereas the gain length of a conventional diode is 200 to 500 *m*m (the length of the structure), the gain length in a VCSEL is the length of the active layer. Resonator optics are fabricated above and below the semiconductor crystal. With a short active layer and low gain, cavity optics must be fabricated from multiple layers of dielectrics alternating quarter-wavelength-thick layers of high- and low-index-of-refraction materials for high reflectivity. Current in the device flows along the optical axis through electrodes on the top and bottom of the device instead of perpendicular to it. These electrodes can be fabricated so that current flows through the two contact layers close to the junction, that offering a lower electrical resistance.

The output from the VCSEL is preferred for coupling to a fiber since emission occurs in the form of a circular beam which is easily focused, as opposed to the elliptical output beam from an edge emitter.

Distributed feedback laser:

A distributed-feedback laser is a laser where the whole resonator consists of a periodic structure, which acts as a distributed reflector in the wavelength range of laser action, and contains a gain medium. Typically, the periodic structure is made with a phase shift in its middle. This structure is essentially the direct concatenation of two Bragg gratings with optical gain within the gratings. This periodic structure reflects partially at each interface between the materials of differing refractive index, so optical feedback is distributed throughout the cavity [hence the term distributed feedback (DFB)]. The grating, again a corrugated surface composed of dielectric materials, is fabricated into the structure of the diode itself and is distributed over a considerably longer length than a traditional grating stretching over the entire length of the device.

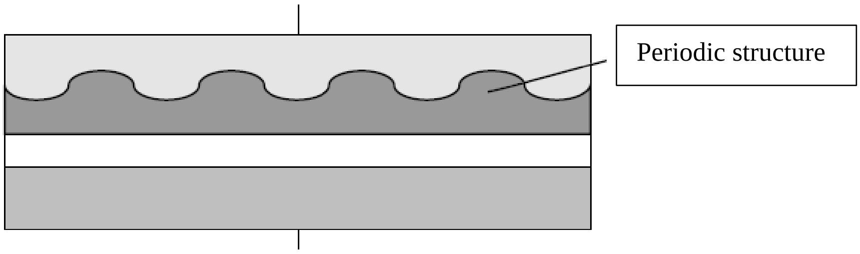


Fig. 2.8 DFB Laser

References:

1. Dennis Roddy, “Satellite communication,” 4th Edition, McGraw Hills.
2. Gerd Keiser, “Optical fiber communication” 2nd Edition, McGraw Hills, 1991.
3. Mark Csele, “Fundamental of light source and Laser” John Willy & Sons Int Publications, 2004.
4. Cees Ronda (Edited), “Luminescence from theory to applications” Willy VCH Verlag GMBH & KGaA, 2008.
5. http://vlab.amrita.edu/?sub=3&brch=269&sim=1371&cnt=1, access on 31-03-2020.
6. https://www.tutorialspoint.com/satellite_communication/index.htm