UNIT I

OVERVIEW OF OPTICAL FIBER COMMUNICATION: INTRODUCTION

Introduction

Fiber-optic communication is a method of transmitting information from one place to another by sending pulses of light through an optical fiber. The light forms an electromagnetic carrier wave that is modulated to carry information.[1] Fiber is preferred over electrical cabling when high bandwidth, long distance, or immunity to electromagnetic interference are required. This type of communication can transmit voice, video, and telemetry through local area networks, computer networks, or across long distances.

Optical fiber is used by many telecommunications companies to transmit telephone signals, Internet communication, and cable television signals. Researchers at Bell Labs have reached internet speeds of over 100 peta bit ×kilometer per second using fiber-optic communication.

The process of communicating using fiber-optics involves the following basic steps:

1. creating the optical signal involving the use of a transmitter, usually from an electrical signal
2. relaying the signal along the fiber, ensuring that the signal does not become too distorted or weak
3. receiving the optical signal
4. converting it into an electrical signal

Historical Development

First developed in the 1970s, fiber-optics have revolutionized the telecommunications industry and have played a major role in the advent of the Information Age. Because of its advantages over electrical transmission, optical fibers have largely replaced copper wire communications in core networks in the developed world.

In 1880 Alexander Graham Bell and his assistant Charles Sumner Tainter created a very early precursor to fiber-optic communications, the Photophone, at Bell's newly established Volta Laboratory in Washington, D.C. Bell considered it his most important invention. The device allowed for the transmission of sound on a beam of light. On June 3, 1880, Bell conducted the world's first wireless telephone transmission between two buildings, some 213 meters apart.[4][5] Due to its use of an atmospheric transmission medium, the Photophone would not prove practical until advances in laser and optical fiber technologies permitted the secure transport of light. The Photophone's first practical use came in military communication systems many decades later.

In 1954 Harold Hopkins and Narinder Singh Kapany showed that rolled fiber glass allowed light to be transmitted. Initially it was considered that the light can traverse in only straight medium. Jun-ichi Nishizawa, a Japanese scientist at Tohoku University, proposed the use of optical fibers for communications in 1963. Nishizawa invented the PIN diode and the static induction transistor, both of which contributed to the development of optical fiber communications.

In 1966 Charles K. Kao and George Hockham at STC Laboratories (STL) showed that the losses of 1,000 dB/km in existing glass (compared to 5–10 dB/km in coaxial cable) were due to contaminants which could potentially be removed.

Optical fiber was successfully developed in 1970 by Corning Glass Works, with attenuation low enough for communication purposes (about 20 dB/km) and at the same time GaAs semiconductor lasers were developed that were compact and therefore suitable for transmitting light through fiber optic cables for long distances.

In 1973, Optelecom, Inc., co-founded by the inventor of the laser, Gordon Gould, received a contract from APA for the first optical communication systems. Developed for Army Missile Command in Huntsville, Alabama, it was a laser on the ground and a spout of optical fiber played out by missile to transmit a modulated signal over five kilometers.

After a period of research starting from 1975, the first commercial fiber-optic communications system was developed which operated at a wavelength around 0.8 µm and used GaAs semiconductor lasers. This first-generation system operated at a bit rate of 45 Mbit/s with repeater spacing of up to 10 km. Soon on 22 April 1977, General Telephone and Electronics sent the first live telephone traffic through fiber optics at a 6 Mbit/s throughput in Long Beach, California.

In October 1973, Corning Glass signed a development contract with CSELT and Pirelli aimed to test fiber optics in an urban environment: in September 1977, the second cable in this test series, named COS-2, was experimentally deployed in two lines (9 km) in Turin, for the first time in a big city, at a speed of 140 Mbit/s.

The second generation of fiber-optic communication was developed for commercial use in the early 1980s, operated at 1.3 µm and used InGaAsP semiconductor lasers. These early systems were initially limited by multi mode fiber dispersion, and in 1981 the single-mode fiber was revealed to greatly improve system performance, however practical connectors capable of working with single mode fiber proved difficult to develop. Canadian service provider SaskTel had completed construction of what was then the world's longest commercial fiber optic network, which covered 3,268 km (2,031 mi) and linked 52 communities.[11] By 1987, these systems were operating at bit rates of up to 1.7 Gb/s with repeater spacing up to 50 km (31 mi).

The first transatlantic telephone cable to use optical fiber was TAT-8, based on Desurvire optimised laser amplification technology. It went into operation in 1988.

Third-generation fiber-optic systems operated at 1.55 µm and had losses of about 0.2 dB/km. This development was spurred by the discovery of Indium gallium arsenide and the development of the Indium Gallium Arsenide photodiode by Pearsall. Engineers overcame earlier difficulties with pulse-spreading at that wavelength using conventional InGaAsP semiconductor lasers. Scientists overcame this difficulty by using dispersion-shifted fibers designed to have minimal dispersion at 1.55 µm or by limiting the laser spectrum to a single longitudinal mode.

These developments eventually allowed third-generation systems to operate commercially at 2.5 Gbit/s with repeater spacing in excess of 100 km (62 mi).

The fourth generation of fiber-optic communication systems used optical amplification to reduce the need for repeaters and wavelength-division multiplexing to increase data capacity. These two improvements caused a revolution that resulted in the doubling of system capacity every six months starting in 1992 until a bit rate of 10 Tb/s was reached by 2001. In 2006 a bit-rate of 14 Tbit/s was reached over a single 160 km (99 mi) line using optical amplifiers.

The focus of development for the fifth generation of fiber-optic communications is on extending the wavelength range over which a WDM system can operate. The conventional wavelength window, known as the C band, covers the wavelength range 1.53–1.57 µm, and *dry fiber* has a low-loss window promising an extension of that range to 1.30–1.65 µm. Other developments include the concept of "optical solutions", pulses that preserve their shape by counteracting the effects of dispersion with the nonlinear effects of the fiber by using pulses of a specific shape.

In the late 1990s through 2000, industry promoters, and research companies such as KMI, and RHK predicted massive increases in demand for communications bandwidth due to increased use of the Internet, and commercialization of various bandwidth-intensive consumer services, such as video on demand. Internet protocol data traffic was increasing exponentially, at a faster rate than integrated circuit complexity had increased under Moore's Law. From the bust of the dot-com bubble through 2006, however, the main trend in the industry has been consolidation of firms and offshoring of manufacturing to reduce costs. Companies such as Verizon and AT&T have taken advantage of fiber-optic communications to deliver a variety of high-throughput data and broadband services to consumers' homes.

Advantages of Fiber Optic Transmission

Optical fibers have largely replaced copper wire communications in core networks in the developed world, because of its advantages over electrical transmission. Here are the main advantages of fiber optic transmission.

Extremely High Bandwidth: No other cable-based data transmission medium offers the bandwidth that fiber does. The volume of data that fiber optic cables transmit per unit time is far great than copper cables.

Longer Distance: in fiber optic transmission, optical cables are capable of providing low power loss, which enables signals can be transmitted to a longer distance than copper cables.

Resistance to Electromagnetic Interference: in practical cable deployment, it's inevitable to meet environments like power substations, heating, ventilating and other industrial sources of interference. However, fiber has a very low rate of bit error (10 EXP-13), as a result of fiber being so resistant to electromagnetic interference. Fiber optic transmission is virtually noise free.

Low Security Risk: the growth of the fiber optic communication market is mainly driven by increasing awareness about data security concerns and use of the alternative raw material. Data or signals are transmitted via light in fiber optic transmission. Therefore there is no way to detect the data being transmitted by "listening in" to the electromagnetic energy "leaking" through the cable, which ensures the absolute security of information.

Small Size: fiber optic cable has a very small diameter. For instance, the cable diameter of a single OM3 multimode fiber is about 2mm, which is smaller than that of coaxial copper cable. Small size saves more space in fiber optic transmission.

Light Weight: fiber optic cables are made of glass or plastic, and they are thinner than copper cables. These make them lighter and easy to install.

Easy to Accommodate Increasing Bandwidth: with the use of fiber optic cable, new equipment can be added to existing cable infrastructure. Because optical cable can provide vastly expanded capacity over the originally laid cable. And WDM (wavelength division multiplexing) technology, including CWDM and DWDM, enables fiber cables the ability to accommodate more bandwidth.

Disadvantages of Fiber Optic Transmission

Though fiber optic transmission brings lots of convenience, its disadvantages also cannot be ignored. **Fragility:** usually optical fiber cables are made of glass, which lends to they are more fragile than electrical wires. In addition, glass can be affected by various chemicals including hydrogen gas (a problem in underwater cables), making them need more cares when deployed under ground.

Difficult to Install: it's not easy to splice fiber optic cable. And if you bend them too much, they will break. And fiber cable is highly susceptible to becoming cut or damaged during installation or construction activities. All these make it difficult to install.

Attenuation & Dispersion: as transmission distance getting longer, light will be attenuated and dispersed, which requires extra optical components like EDFA to be added.

Cost Is Higher Than Copper Cable: despite the fact that fiber optic installation costs are dropping by as much as 60% a year, installing fiber optic cabling is still relatively higher than copper cables. Because copper cable installation does not need extra care like fiber cables. However, optical fiber is still moving into the local loop, and through technologies such as FTTx (fiber to the home, premises, etc.) and PONs (passive optical networks), enabling subscriber and end user broadband access.

Special Equipment Is Often Required: to ensure the quality of fiber optic transmission, some special equipment is needed. For example, equipment such as OTDR (optical time-domain reflectometry) is required and expensive, specialized optical test equipment such as optical probes and power meter are needed at most fiber endpoints to properly provide testing of optical fiber.

Applications of Optical Fiber Communications

Fiber optic cables find many uses in a wide variety of industries and applications. Some uses of fiber optic cables include:

- **Medical**

  Used as light guides, imaging tools and also as lasers for surgeries

- **Defense/Government**

  Used as hydrophones for seismic waves and SONAR , as wiring in aircraft, submarines and other vehicles and also for field networking

- **Data Storage**

  Used for data transmission

- **Telecommunications**

  Fiber is laid and used for transmitting and receiving purposes

- **Networking**

  Used to connect users and servers in a variety of network settings and help increase the speed and accuracy of data transmission

- **Industrial/Commercial**

  Used for imaging in hard to reach areas, as wiring where EMI is an issue, as sensory devices to make temperature, pressure and other measurements, and as wiring in automobiles and in industrial settings

- **Broadcast/CATV**

  Broadcast/cable companies are using fiber optic cables for wiring CATV, HDTV, internet, video on-demand and other applications

  Fiber optic cables are used for lighting and imaging and as sensors to measure and monitor a vast array of variables. Fiber optic cables are also used in research and development and testing across all the above mentioned industries

The optical fibers have many applications. Some of them are as follows –

- Used in telephone systems
- Used in sub-marine cable networks
- Used in data link for computer networks, CATV Systems
- Used in CCTV surveillance cameras
- Used for connecting fire, police, and other emergency services.
- Used in hospitals, schools, and traffic management systems.
- They have many industrial uses and also used for in heavy duty constructions.

Block Diagram of Optical Fiber Communication System

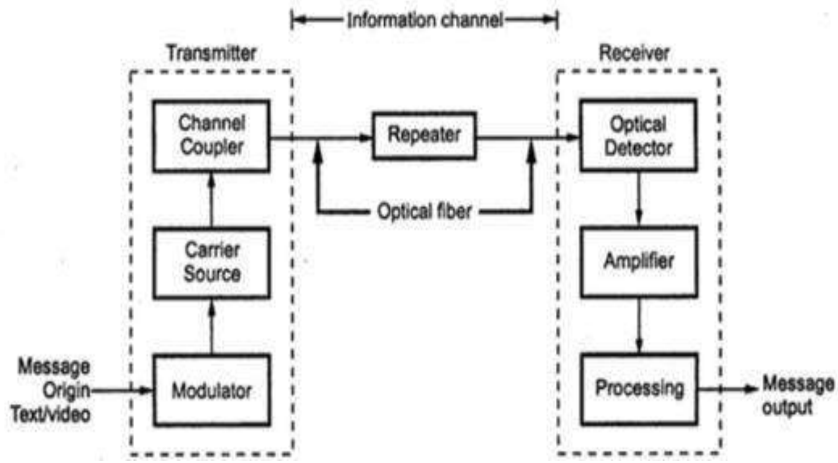


Fig 1: Block Diagram of Optical Fiber Communication System

Message origin:

Generally message origin is from a transducer that converts a non-electrical message into an electrical signal. Common examples include microphones for converting sound waves into currents and video (TV) cameras for converting images into current. For data transfer between computers, the message is already in electrical form.

Modulator:

The modulator has two main functions.

1) It converts the electrical message into proper format.

2) It impresses this signal onto the wave generated by the carrier source.

wo distinct categories of modulation are used i.e. analog modulation and digital modulation.

Carrier source:

. Carrier source generates the wave on which the information is transmitted. This wave is called the carrier. For fiber optic system, a laser diode (LD) or a light emitting diode (LED) is used. They can be called as optic oscillators, they provide stable, single frequency waves with sufficient power for long distance propagation.

Channel coupler:

. Coupler feeds the power into information channel. For an atmospheric optic system, the channel coupler is a lens used for collimating the light emitted by the source and directing this light towards the receiver. The coupler must efficiently transfer the modulated light beam from the source to the optic fiber. The channel coupler design is an important part of fiber system because of possibility of high losses.

Information channel:

. The information channel is the path between the transmitter and receiver. In fiber optic communications, a glass or plastic fiber is the channel. Desirable characteristics of the information channel include low attenuation and large light acceptance cone angle. Optical amplifiers boost the power levels of weak signals. Amplifiers are needed in very long links to provide sufficient power to the receiver. Repeaters can be used only for digital systems. They convert weak and distorted optical signals to electrical ones and then regenerate the original digital pulse trains for further transmission.

. Another important property of the information channel is the propagation time of the waves travelling along it. A signal propagating along a fiber normally contains a range of fiber optic frequencies and divides its power along several ray paths. This results in a distortion of the propagation signal. In a digital system, this distortion appears as a spreading and deforming of the pulses. The spreading is so great that adjacent pulses begin to overlap and become unrecognizable as separate bits of information.

Optical detector:

. The information begin transmitted is detected by detector. In the fiber system the optic wave is converted into an electric current by a photodetector. The current developed by the detector is proportional to the power in the incident optic wave. Detector output current contains the transmitted information. This detector output is then filtered to remove the constant bias and then amplified.

. The important properties of photodetectors are small size, economy, long life, low power consumption, high sensitivity to optic signals and fast response to quick variations in the optic power.

. Signal processing includes filtering, amplification. Proper filtering maximizes the ratio of signal to unwanted power. For a digital syst5em decision circuit is an additional block. The bit error rate (BER) should be very small for quality communications.

Signal processing:

. Signal processing includes filtering, amplification. Proper filtering maximizes the ratio of signal to unwanted power. For a digital syst5em decision circuit is an additional block. The bit error rate (BER) should be very small for quality communications.

Message output:

. The electrical form of the message emerging from the signal processor is transformed into a sound wave or visual image. Sometimes these signals are directly usable when computers or other machines are connected through a fiber system.

Electromagnetic Spectrum

The radio waves and light are electromagnetic waves. The rate at which they alternate in
polarity is called their frequency (f) measured in hertz (Hz). The speed of electromagnetic wave (c) in free space is approximately 3 x 108 m/sec. The distance travelled during each cycle is called as wavelength (λ)

In fiber optics, it is more convenient to use the wavelength of light instead of the frequency with light frequencies; wavelength is often stated in microns or nanometers.

1 micron (μ) = 1

Micrometre (1 x 10-6) 1 nano (n) = 10-9 meter

Fiber optics uses visible and infrared light. Infrared light covers a fairly wide range of wavelengths and is generally used for all fiber optic communications. Visible light is normally used for very short range transmission using a plastic fiber.

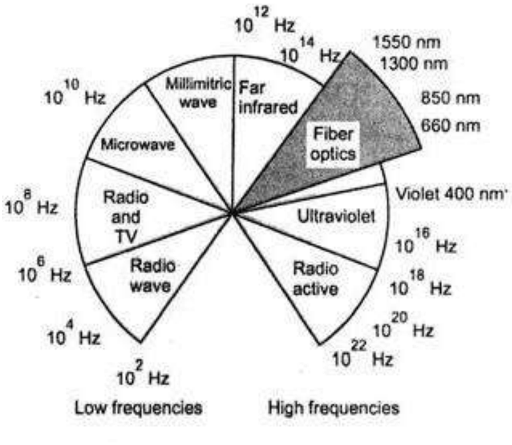
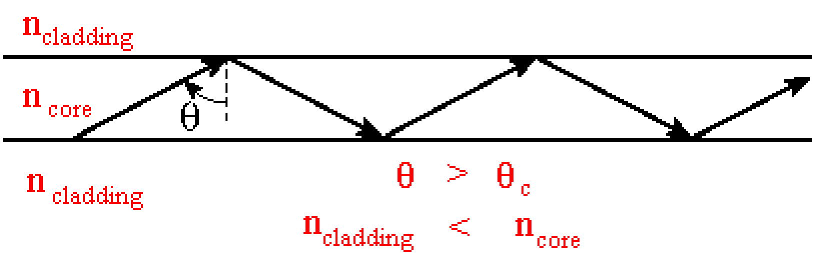


Fig 2: Electromagnetic Spectrum

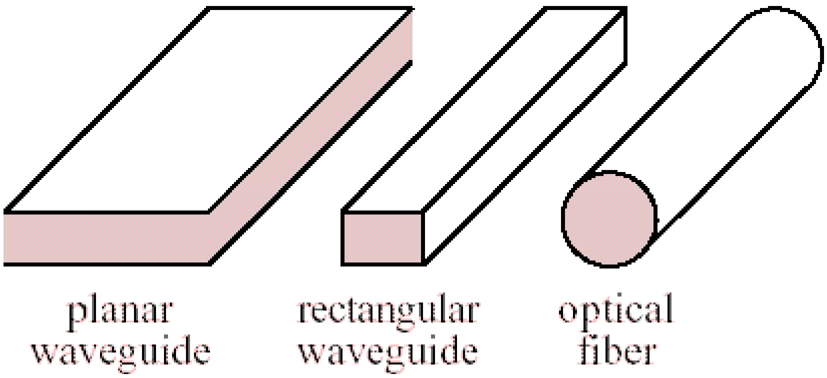
Optical Fiber Waveguides

In free space light ravels as its maximum possible speed i.e. 3 x 108 m/s or 186 x 103 miles/sec. When light travels through a material it exhibits certain behavior explained by laws of reflection, refraction.

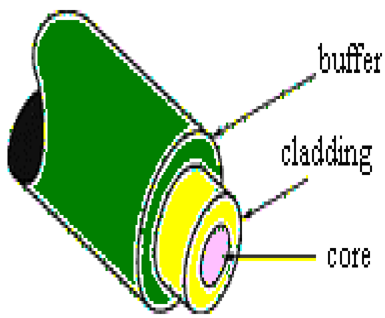
An optical wave guide is a structure that "guides" a light wave by constraining it to travel along a certain desired path. If the transverse dimensions of the guide are much larger than the wavelength of the guided light, then we can explain how the optical waveguide works using geometrical optics and total internal reflection.



A wave guide traps light by surrounding a guiding region, called the core, made from a material with index of refraction n_{core}, with a material called the cladding, made from a material with index of refraction $\mathbf{n_{cladding} < n_{core}}$. Light entering is trapped as long as $\mathbf{\sin\theta > n_{cladding}/n_{ncore}}$.

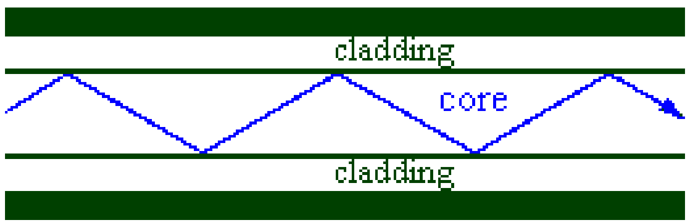


Light can be guided by planar or rectangular wave guides, or by optical fibers.An optical fiber consists of three concentric elements, the core, the cladding and the outer coating, often called the buffer. The core is usually made of glass or plastic. The core is the light-carrying portion of the fiber. The cladding surrounds the core. The cladding is made of a material with a slightly lower index of refraction than the core. This difference in the indices causes total internal reflection to occur at the core-cladding boundary along the length of the fiber. Light is transmitted down the fiber and does not escape through the sides of the fiber.

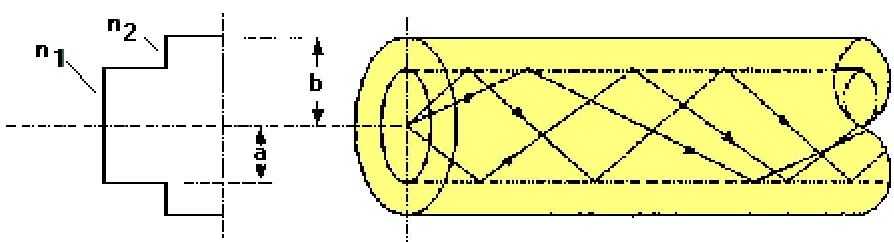


- Fiber Optic Core:
    - the inner light-carrying member with a high index of refraction.
- Cladding:
    - the middle layer, which serves to confine the light to the core. It has a lower index of refraction.

- Buffer:
    - the outer layer, which serves as a "shock absorber" to protect the core and cladding from damage. The coating usually comprises one or more coats of a plastic material to protect the fiber from the physical environment. Sometimes metallic sheaths are added to the coating for further physical protection.



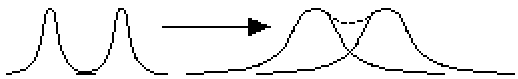
- Light injected into the fiber optic core and striking the core-to-cladding interface at an angle greater than the critical angle is reflected back into the core. Since the angles of incidence and reflection are equal, the light ray continues to zigzag down the length of the fiber. The light is trapped within the core. Light striking the interface at less than the critical angle passes into the cladding and is lost.



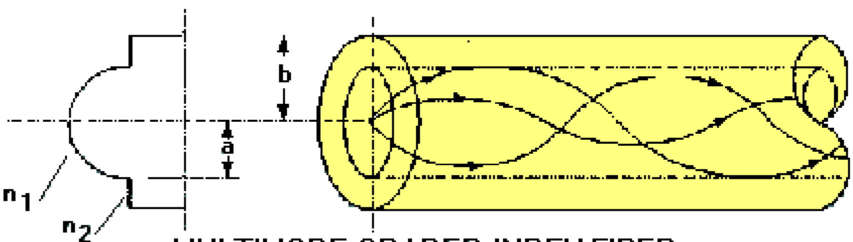
MULTIMODE STEP-INDEX FIBER

- Fibers for which the refractive index of the core is a constant and the index changes abruptly at the core-cladding interface are called step-index fibers. Step-index fibers are available with core diameters of 100 mm to 1000 mm. They are well suited to applications requiring high-power densities, such as delivering laser power for medical and industrial applications.

- Multimode step-index fibers trap light with many different entrance angles, each mode in a step-index multimode fiber is associated with a different entrance angle. Each mode therefore travels along a different path through the fiber. Different propagating modes have different velocities. As an optical pulse travels down a multimode fiber, the pulse begins to spread. Pulses that enter well separated from each other will eventually overlap each other. This limits the distance over which the fiber can transport data. Multimode step-index fibers are not well suited for data transport and communications.

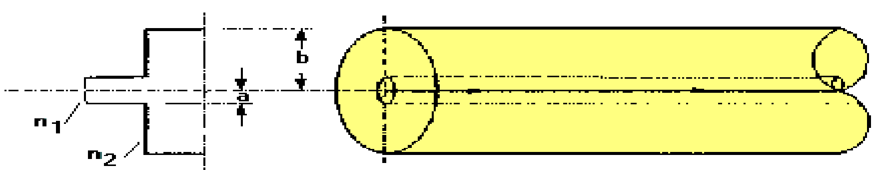
- 

- In a multimode graded-index fiber the core has an index of refraction that decreases as the radial distance from the center of the core increases. As a result, the light travels faster near the edge of the core than near the center. Different modes therefore travel in curved paths with nearly equal travel times. This greatly reduces the spreading of optical pulses.



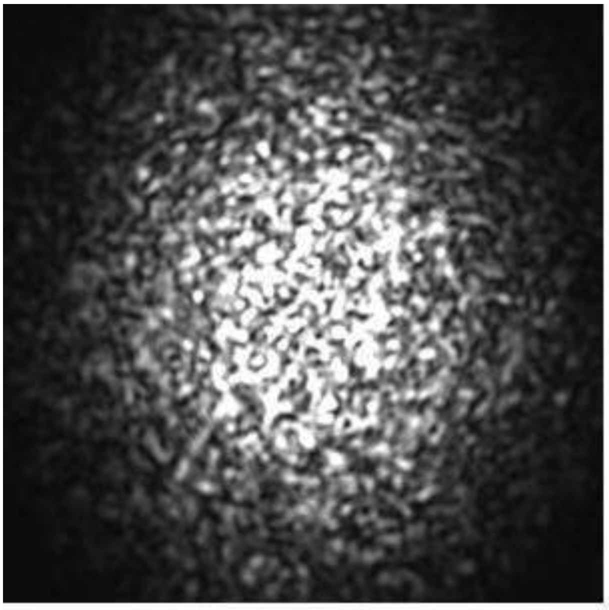
MULTIMODE GRADED-INDEX FIBER

-
- A **single mode fiber** only allows light to propagate down its center and there are no longer different velocities for different modes. A single mode fiber is much thinner than a multimode fiber and can no longer be analyzed using geometrical optics. Typical core diameters are between 5 mm and 10 mm.

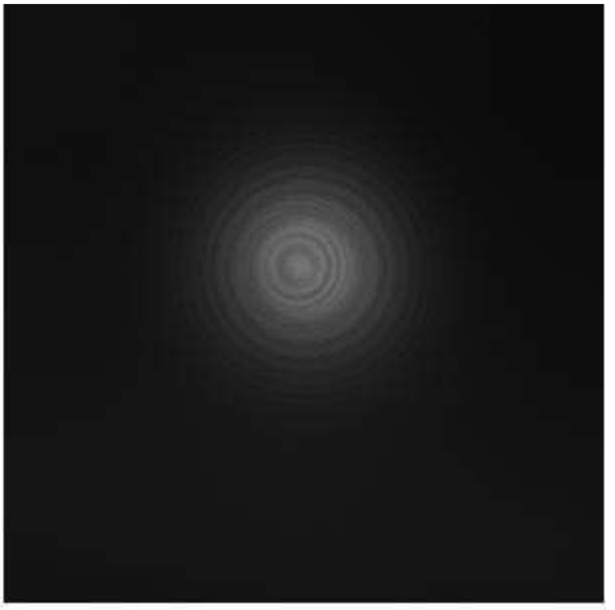


SINGLE MODE STEP-INDEX FIBER

When laser light is coupled into a fiber, the distribution of the light emerging from the other end reveals if the fiber is a multimode or single mode fiber.



Light emerging from a multi-mode fiber

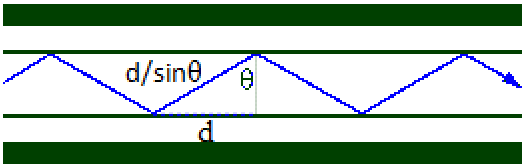


Light emerging from a single-mode fiber

Optical fibers are used widely in the medical field for diagnoses and treatment. Optical fibers can be bundled into flexible strands, which can be inserted into blood vessels, lungs and other parts of the body. An Endoscope is a medical tool carrying two bundles of optic fibers inside one long tube. One bundle directs light at the tissue being tested, while the other bundle carries light reflected from the tissue, producing a detailed image. Endoscopes can be designed to look at regions of the human body, such as the knees, or other joints in the body

Problem:

In a step-index fiber in the ray approximation, the ray propagating along the axis of the fiber has the shortest route, while the ray incident at the critical angle has the longest route. Determine the difference in travel time (in ns/km) for the modes defined by those two rays for a fiber with $n_{core} = 1.5$ and $n_{cladding} = 1.485$.



Solution:

If a ray propagating along the axis of the fiber travels a distance d, then a ray incident at the critical angle θ_c travels a distance $L = d/\sin\theta_c$.

The respective travel times are $t_d = dn_{core}/c$ and $t_L = dn_{core}/(\sin\theta_c\, c)$.

$\sin\theta_c = n_{cladding}/n_{core}$.

$\theta_c = 81.9$ deg.

For d = 1000 m we have $t_d = 5000$ ns and $t_L = 5050.51$ ns.

The difference in travel time is therefore 50.51 ns/km.

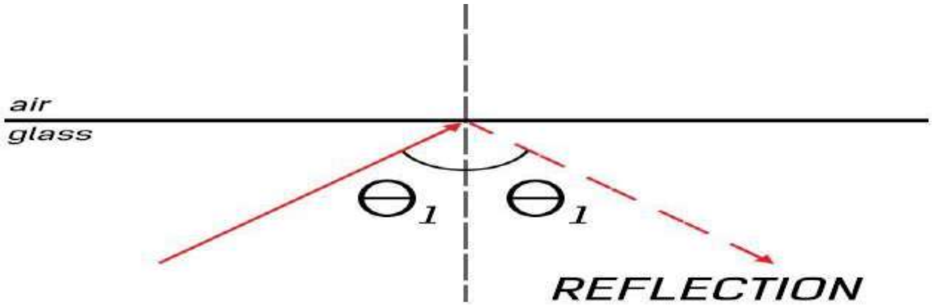
Ray theory

The phenomenon of splitting of white light into its constituents is known as dispersion. The concepts of reflection and refraction of light are based on a theory known as Ray theory or geometric optics, where light waves are considered as waves and represented with simple geometric lines or rays.

The basic laws of ray theory/geometric optics

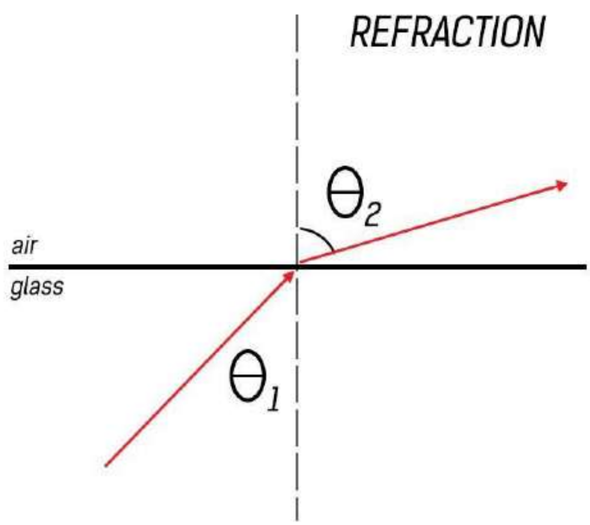
- In a homogeneous medium, light rays are straight lines.
- Light may be absorbed or reflected
- Reflected ray lies in the plane of incidence and angle of incidence will be equal to the angle of reflection.

- At the boundary between two media of different refractive indices, the refracted ray will lie in the plane of incidence. Snell's Law will give the relationship between the angles of incidence and refraction.



Reflection depends on the type of surface on which light is incident. An essential condition for reflection to occur with glossy surfaces is that the angle made by the incident ray of light with the normal at the point of contact should be equal to the angle of reflection with that normal.

The *images* produced from this reflection have different properties according to the shape of the surface. For example, for a flat mirror, the image produced is upright, has the same size as that of the object and is equally distanced from the surface of the mirror as the real object. However, the properties of a parabolic mirror are different and so on.



Refraction is the bending of light in a particular medium due to the speed of light in that medium. The speed of light in any medium can be given by $v = \frac{c}{n}$

$$\text{Refractive index n} = \frac{\text{Speed of light in air}}{\text{Speed of light in medium}} = \frac{c}{v}$$

The refractive index for vacuum and air os 1.0 for water it is 1.3 and for glass refractive index is 1.5. Here n is the **refractive index** of that medium. When a ray of light is incident at the interface of two media with different refractive indices, it will bend either towards or away from the normal depending on the refractive indices of the media.

According to **Snell's law**, refraction can be represented as

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2)$$

n_1 = refractive index of first medium

θ_1 = angle of incidence

n_2 = refractive index of second medium

θ_2 = angle of refraction

For $n_1 > n_2$, θ_2 is always greater than θ_1. **Or** to put it in different words, light moving from a medium of high refractive index (glass) to a medium of lower refractive index (air) will move away from the normal.

Total internal reflection

To consider the propagation of light within an optical fiber utilizing the ray theory model it is necessary to take account of the refractive index of the dielectric medium. Optical materials are characterized by their index of refraction, referred to as n.The refractive index of a medium is defined as the ratio of the velocity of light in a vacuum to the velocity of light in the medium.

When a beam of light passes from one material to another with a different index of refraction, the beam is bent (or refracted) at the interface (Figure 2).

$$n_I \sin I = n_R \sin R$$

where n_I and n_R are the indices of refraction of the materials through which the beam is refracted and *I* and *R* are the angles of incidence and refraction of the beam. If the angle of incidence is greater than the critical angle for the interface (typically about 82° for optical fibers), the light is reflected back into the incident medium without loss by a process known as total internal reflection (Figure 3).

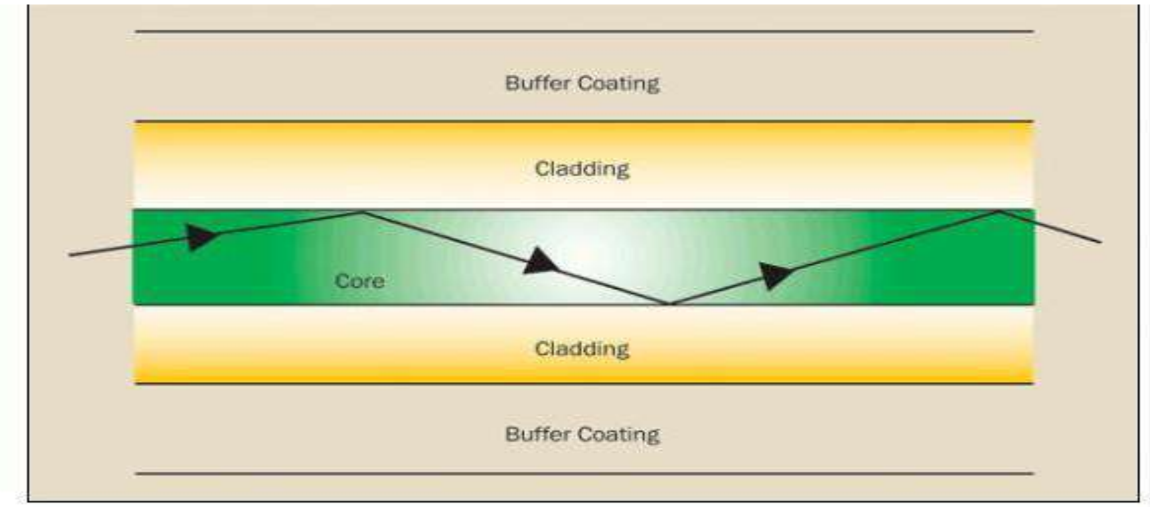


Figure 3. *Total internal reflection allows light to remain inside the core of the fiber.*

Refraction is described by Snell's law:

A ray of light travels more slowly in an optically dense medium than in one that is less dense, and the refractive index gives a measure of this effect. When a ray is incident on the interface between two dielectrics of differing refractive indices (e.g. glass–air), refraction occurs, as illustrated in Figure 1.2(a). It may be observed that the ray approaching the interface is propagating in a dielectric of refractive index *n* and is at an angle ϕ to the normal at the surface of the interface.

If the dielectric on the other side of the interface has a refractive index *n* which is less than *n*1, then the refraction is such that the ray path in this lower index medium is at an angle to the normal, where is greater than . The angles of incidence and refraction are related to each other and to the refractive indices of the dielectrics by Snell's law of refraction, which states that:

$$n_1 \sin \phi_1 = n_2 \sin \phi_2$$

Or

$$\frac{\sin \phi_1}{\sin \phi_2} = \frac{n_2}{n_1}$$

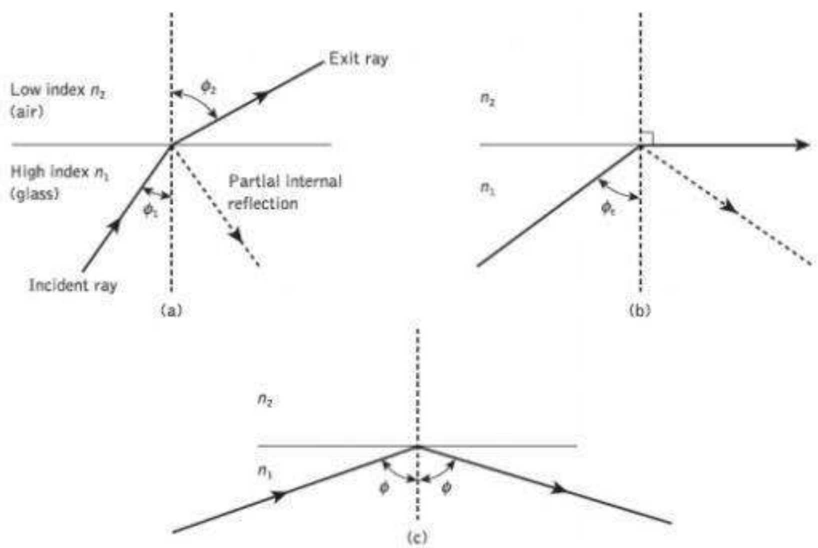


Figure 1.2 Light rays incident on a high to low refractive index interface (e.g. glass air): (a) refraction; (b) the limiting case of refraction showing the critical ray at an angle φ_c (c) total internal reflection where $\varphi > \varphi_c$

It may also be observed in Figure 1.2(a) that a small amount of light is reflected back into the originating dielectric medium (partial internal reflection). As *n* is greater than *n*, the angle of refraction is always greater than the angle of incidence. Thus when the angle of refraction is 90° and the refracted ray emerges parallel to the interface between the dielectrics, the angle of incidence must be less than 90°.

This is the limiting case of refraction and the angle of incidence is now known as the critical angle ϕc, as shown in Figure 1.2(b). From Eq. (1.1) the value of the critical angle is given by

$$\sin \phi_c = \frac{n_2}{n_1}$$

At angles of incidence greater than the critical angle the light is reflected back into the originating dielectric medium (total internal reflection) with high efficiency (around 99.9%). Hence, it may be observed in Figure 1.2(c) that total internal reflection occurs at the inter- face between two dielectrics of differing refractive indices when light is incident on the dielectric of lower index from the dielectric of higher index, and the angle of incidence of the ray exceeds the critical value. This is the mechanism by which light at a sufficiently shallow angle (less than 90° − may be considered to propagate down an optical fiber with low loss.

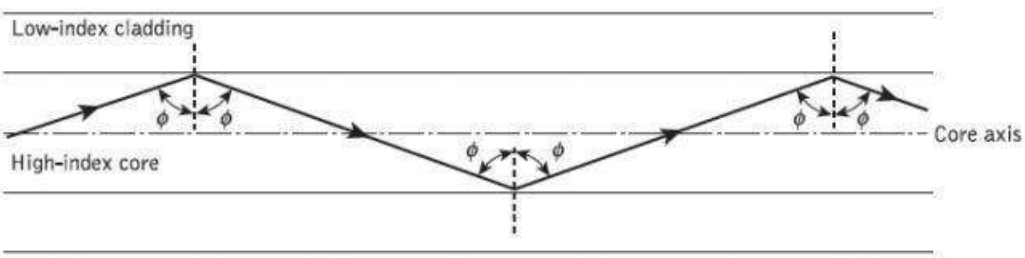


Figure 1.3 The transmission of a light ray in a perfect optical fiber

Figure 1.3 illustrates the transmission of a light ray in an optical fiber via a series of total internal reflections at the interface of the silica core and the slightly lower refractive index silica cladding. The ray has an angle of incidence ϕ at the interface which is greater than the critical angle and is reflected at the same angle to the normal.

The light ray shown in Figure 1.3 is known as a meridional ray as it passes through the axis of the fiber core. This type of ray is the simplest to describe and is generally used when illustrating the fundamental transmission properties of optical fibers. It must also be noted that the light transmission illustrated in Figure 1.3 assumes a perfect fiber, and that any discontinuities or imperfections at the core–cladding interface would probably result in refraction rather than total internal reflection, with the subsequent loss of the light ray into the cladding.

Critical Angle

When the angle of incidence (ϕ_1) is progressively increased, there will be progressive increase of refractive angle (ϕ_2). At some condition (ϕ_1) the refractive angle (ϕ_2) becomes 90° to the normal. When this happens the refracted light ray travels along the interface. The angle of incidence (ϕ_1) at the point at which the refractive angle (ϕ_1) becomes 90° is called the critical angle. It is denoted by ϕ_c.

The **critical angle** is defined as the minimum angle of incidence (ϕ_1) at which the ray strikes the interface of two media and causes an angle of refraction (ϕ_2) equal to 90°. Fig 1.6.5 shows critical angle refraction. When the angle of refraction is 90 degree to the normal the refracted ray is parallel to the interface between the two media.

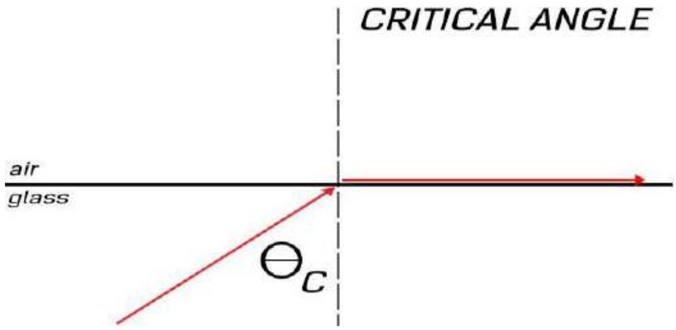
Hence at critical angle $\phi_1 = \phi_c$ and $\phi_2 = 90^\circ$

Using Snell's law: $n_1 \sin \phi_1 = n_2 \sin \phi_2$

$$\sin \phi_c = \frac{n_2}{n_1} \sin 90^\circ$$

$$\because \quad \sin 90^\circ = 1$$

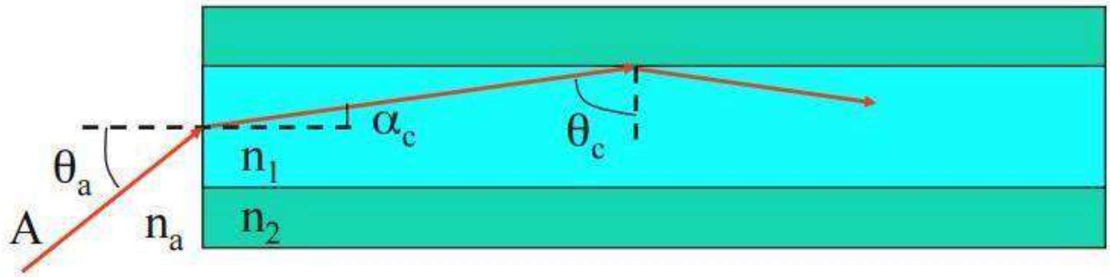
$$\boxed{\text{Critical angle } \phi_c = \sin^{-1}\left(\frac{n_2}{n_1}\right)}$$



It is important to know about this property because reflection is also possible even if the surfaces are not reflective. If the *angle of incidence is greater than the critical angle* for a given setting, the resulting type of reflection is called **Total Internal Reflection**, and it is the basis of Optical Fiber Communication.

Acceptance angle

In an optical fiber, a light ray undergoes its *first refraction* at the air-core interface. The angle at which this refraction occurs is crucial because this particular angle will dictate whether the subsequent *internal* reflections will follow the principle of Total Internal Reflection. This angle, at which the light ray first encounters the core of an optical fiber is called Acceptance angle.

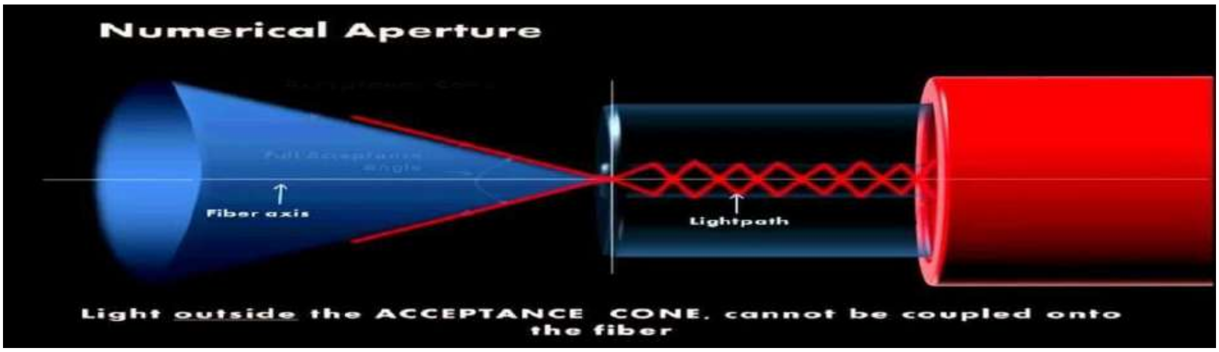


The objective is to have [latex] \theta_{c}[/latex] greater than the critical angle for this particular setting. As you can notice, θ_c depends on the orientation of the refracted ray at the input of the optical fiber. This in turn depends on θ_a, the acceptance angle.

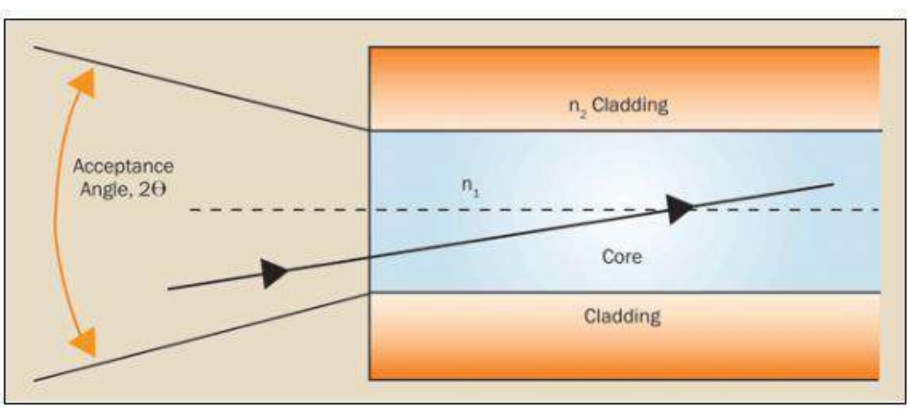
The acceptance angle can be calculated with the help of the formula below.

Numerical Aperture

Numerical Aperture is a characteristic of any optical system. For example, photo-detector, optical fiber, lenses etc. are all optical systems. Numerical aperture is the ability of the optical system to collect all of the light incident on it, in one area.



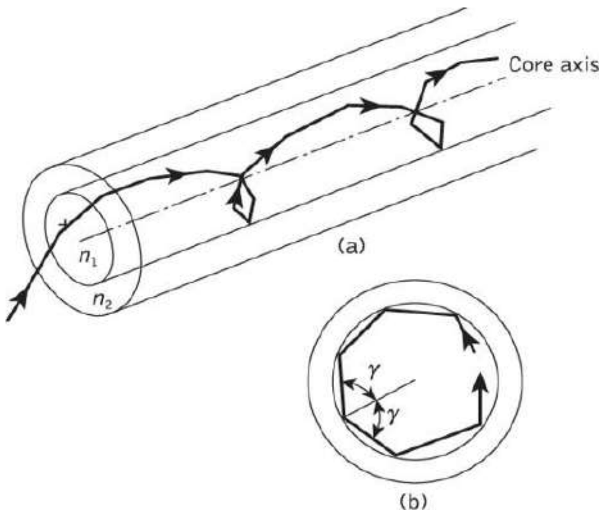
The blue cone is known as the cone of acceptance. As you can see it is dependent on the Acceptance Angle of the optical fiber. Light waves within the acceptance cone can be collected in a small area which can then be sent into the optical fiber (Source)



Numerical aperture (NA), shown in above Figure, is the measure of maximum angle at which light rays will enter and be conducted down the fiber. This is represented by the following equation:

$$\mathrm{NA} = \sqrt{(n^2_{core} - n^2_{cladding})} = \sin\theta$$

skew rays: In a multimode optical fiber, a bound ray that travels in a helical path along the fiber and thus (a) is not parallel to the fiber axis, (b) does not lie in a meridional plane, and (c) does not intersect the fiber axis is known as a Skew Ray.



1. Skew rays are rays that travel through an optical fiber without passing through its axis.
2. A possible path of propagation of skew rays is shown in figure. Figure 24, view (a), provides an angled view and view (b) provides a front view.

3. Skew rays are those rays which follow helical path but they are not confined to a single plane. Skew rays are not confined to a particular plane so they cannot be tracked easily. Analyzing the meridional rays is sufficient for the purpose of result, rather than skew rays, because skew rays lead to greater power loss.

4. Skew rays propagate without passing through the center axis of the fiber. The acceptance angle for skew rays is larger than the acceptance angle of meridional rays.

5. Skew rays are often used in the calculation of light acceptance in an optical fiber. The addition of skew rays increases the amount of light capacity of a fiber. In large NA fibers, the increase may be significant.

6. The addition of skew rays also increases the amount of loss in a fiber. Skew rays tend to propagate near the edge of the fiber core. A large portion of the number of skew rays that are trapped in the fiber core are considered to be leaky rays.

7. Leaky rays are predicted to be totally reflected at the core-cladding boundary. However, these rays are partially refracted because of the curved nature of the fiber boundary. Mode theory is also used to describe this type of leaky ray loss.

Cylindrical fiber

1. Modes

When light is guided down a fiber (as microwaves are guided down a waveguide), phase shifts occur at every reflective boundary. There is a finite discrete number of paths down the optical fiber (known as modes) that produce constructive (in phase and therefore additive) phase shifts that reinforce the transmission. Because each mode occurs at a different angle to the fiber axis as the beam travels along the length, each one travels a different length through the fiber from the input to the output. Only one mode, the zero-order mode, travels the length of the fiber without reflections from the sidewalls. This is known as a single-mode fiber. The actual number of modes that can be propagated in a given optical fiber is determined by the wavelength of light and the diameter and index of refraction of the core of the fiber.

The exact solution of Maxwell's equations for a cylindrical homogeneous core dielectric waveguide* involves much algebra and yields a complex result. Although the presentation of this mathematics is beyond the scope of this text, it is useful to consider the resulting modal fields. In common with the planar guide (Section 1.3.2), TE (where $E_z = 0$) and TM (where $H_z = 0$) modes are obtained within the dielectric cylinder. The cylindrical waveguide, however, is bounded in two dimensions rather than one. Thus two integers, l and m, are necessary in order to specify the modes, in contrast to the single integer (m) required for the planar guide.

For the cylindrical waveguide we therefore refer to TE_{lm} and TM_{lm} modes. These modes correspond to meridional rays (see Section 1.2.1) traveling within the fiber. However, hybrid modes where E_z and H_z are nonzero also occur within the cylindrical waveguide.

These modes, which result from skew ray propagation (see Section 1.2.4) within the fiber, are designated HE*lm* and EH*lm* depending upon whether the components of **H** or **E** make the larger contribution to the transverse (to the fiber axis) field. Thus an exact description of the modal fields in a step index fiber proves somewhat complicated.

Fortunately, the analysis may be simplified when considering optical fibers for communication purposes. These fibers satisfy the weakly guiding approximation where the relative index difference Δ1. This corresponds to small grazing angles θ in Eq. (1.34). In fact is usually less than 0.03 (3%) for optical communications fibers. For weakly guiding structures with dominant forward propagation, mode theory gives dominant transverse field components. Hence approximate solutions for the full set of HE, EH, TE and TM modes may be given by two linearly polarized components.

These linearly polarized (LP) modes are not exact modes of the fiber except for the fundamental (lowest order) mode. However, as in weakly guiding fibers is very small, then HE– EH mode pairs occur which have almost identical propagation constants. Such modes are said to be degenerate. The superpositions of these degenerating modes characterized by a common propagation constant correspond to particular LP modes regardless of their HE, EH, TE or TM field configurations. This linear combination of degenerate modes obtained from the exact solution produces a useful simplification in the analysis of weakly guiding fibers.

The relationship between the traditional HE, EH, TE and TM mode designations and the LP*lm* mode designations is shown in Table 1.1. The mode subscripts *l* and *m* are related to the electric field intensity profile for a particular LP mode (see Figure 1.11(d)). There are in general 2*l* field maxima around the circumference of the fiber core and *m* field maxima along a radius vector. Furthermore, it may be observed from Table 1.1 that the notation for labeling the HE and EH modes has changed from that specified for the exact solution in the cylindrical waveguide mentioned previously.

Table 1.1 Correspondence between the lower order in linearly polarized modes and the traditional exact modes from which they are formed

Linearly polarized	*Exact*
LP_{01}	HE_{11}
LP_{11}	HE_{21}, TE_{01}, TM_{01}
LP_{21}	HE_{31}, EH_{11}
LP_{02}	HE_{12}
LP_{31}	HE_{41}, EH_{21}
LP_{12}	HE_{22}, TE_{02}, TM_{02}
LP_{lm}	HE_{2m}, TE_{0m}, TM_{0m}
LP_{lm} ($l \neq 0$ or 1)	$HE_{l+1,m}$, $EH_{l-1,m}$

2. Mode coupling

We have thus far considered the propagation aspects of perfect dielectric waveguides. However, waveguide perturbations such as deviations of the fiber axis from straightness, variations in the core diameter, irregularities at the core–cladding interface and refractive index variations may change the propagation characteristics of the fiber. These will have the effect of coupling energy traveling in one mode to another depending on the specific perturbation. Ray theory aids the understanding of this phenomenon, as shown in Figure 1.13, which illustrates two types of perturbation. It may be observed that in both cases the ray no longer maintains the same angle with the axis. In electromagnetic wave theory this corresponds to a change in the propagating mode for the light. Thus individual modes do not normally propagate throughout the length of the fiber without large energy transfers to adjacent modes, even when the fiber is exceptionally good quality and is not strained or bent by its surroundings. This mode conversion is known as mode coupling or mixing. It is usually analyzed using coupled mode equations which can be obtained directly from Maxwell's equations.

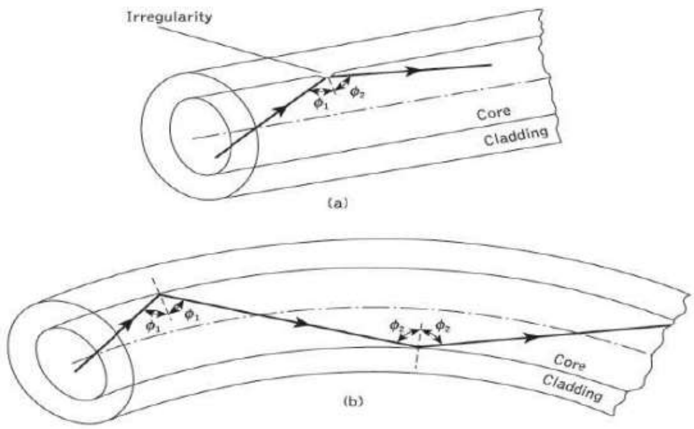


Figure 1.13 *Ray theory illustrations showing two of the possible fiber perturbations which give mode coupling: (a) irregularity at the core–cladding interface; (b) fiber bend*

3. Step index fibers

The optical fiber considered in the preceding sections with a core of constant refractive index *n*1 and a cladding of a slightly lower refractive index *n*2is known as step index fiber. This is because the refractive index profile for this type of fiber makes a step change at the core–cladding interface, as indicated in Figure 1.14, which illustrates the two major types of step index fiber.The refractive index profile may be defined as

$$n(r) = \begin{cases} n_1 & r < a \quad \text{(core)} \\ n_2 & r \geq a \quad \text{(cladding)} \end{cases} \qquad (1.48)$$

in both cases

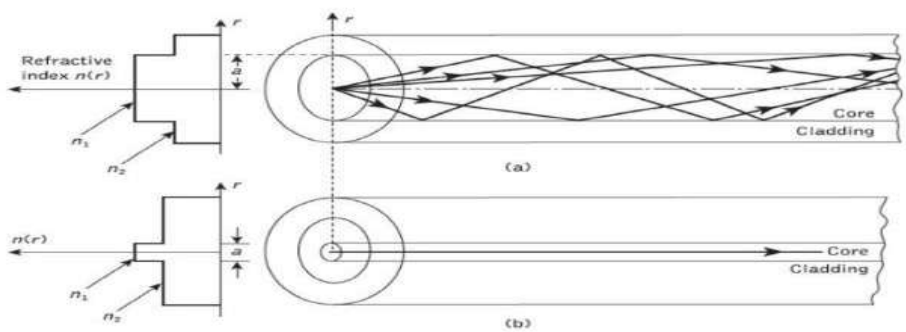


Figure 1.14 The refractive index profile and ray transmission in step index fibers: (a) multimode step index fiber, (b) single-mode step index fiber

Figure 1.14(a) shows a multimode step index fiber with a core diameter of around 50μm or greater, which is large enough to allow the propagation of many modes within the fiber core. This is illustrated in Figure 1.14(a) by the many different possible ray paths through the fiber. Figure 1.14(b) shows a single-mode or monomode step index fiber which allows the propagation of only one transverse electromagnetic mode (typically HE11), and hence the core diameter must be of the order of 2 to 10μm. The propagation of a single mode is illustrated in Figure 1.14(b) as corresponding to a single ray path only (usually shown as the axial ray) through the fiber.

The single-mode step index fiber has the distinct advantage of low intermodal dispersion (broadening of transmitted light pulses), as only one mode is transmitted, whereas with multimode step index fiber considerable dispersion may occur due to the differing group velocities of the propagating modes. This in turn restricts the maximum bandwidth attainable with multimode step index fibers, especially when com- pared with single-mode fibers.

However, for lower bandwidth applications multimode fibers have several advantages over single-mode fibers. These are:

a) The use of spatially incoherent optical sources (e.g. most light-emitting diodes) which cannot be efficiently coupled to single-mode fibers.

b) Larger numerical apertures, as well as core diameters, facilitating easier coupling to optical sources

c) Lower tolerance requirements on fiber connectors

Multimode step index fibers allow the propagation of a finite number of guided modes along the channel. The number of guided modes is dependent upon the physical parameters (i.e. relative refractive index difference, core radius) of the fiber and the wavelengths of the transmitted light which are included in the normalized frequency *V* for the fiber.

Mode propagation does not entirely cease below cutoff. Modes may propagate as unguided or leaky modes which can travel considerable distances along the fiber. Nevertheless, it is the guided modes which are of paramount importance in optical fiber communications as these are confined to the fiber over its full length. that the total number of guided modes or mode volume *M*s for a step index fiber is related to the *V* value for the

fiber by the approximate expression

Which allows an estimate of the number of guided modes propagating in a particular multimode step index fiber.

4. Graded index fibers

Graded index fibers do not have a constant refractive index in the core* but a decreasing core index *n*(*r*) with radial distance from a maximum value of*n*1 at the axis to a constant value *n*2 beyond the core radius *a* in the cladding. This index variation may be represented as:

$$n(r) = \begin{cases} n_1(1 - 2\Delta(r/a)^{\alpha})^{\frac{1}{2}} & r < a \quad \text{(core)} \\ n_1(1 - 2\Delta)^{\frac{1}{2}} = n_2 & r \geq a \quad \text{(cladding)} \end{cases} \qquad (1.50)$$

where is the relative refractive index difference and α is the profile parameter which gives the characteristic refractive index profile of the fiber core. Equation (1.50) which is a convenient method of expressing the refractive index profile of the fiber core as a variation of α, allows representation of the step index profile when α = ∞, a parabolic profile when α = 2 and a triangular profile when α = 1. This range of refractive index profiles is illustrated in Figure 1.15

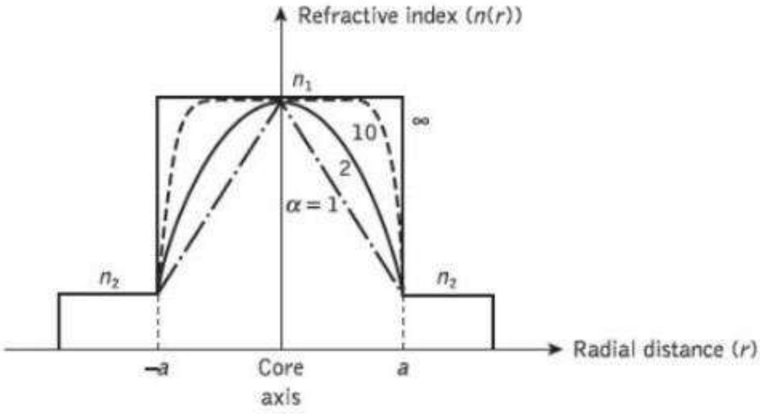


Figure 1.15 Possible fiber refractive index profiles for different values of α□(given in Eq. (1.50)

The graded index profiles which at present produce the best results for multimode optical propagation have a near parabolic refractive index profile core with □~~2. Fibers with such core index profiles are well established and consequently when the term ‘graded index’ is used without qualification it usually refers to a fiber with this profile.

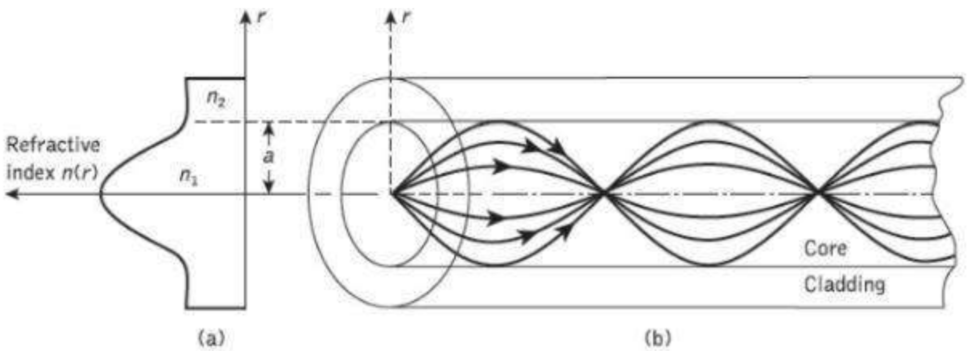


Figure 1.16 The refractive index profile and ray transmission in a multimode graded index fiber

Where r = Radial distance from fiber axis

$$n(r) = \begin{cases} n_1\left(1 - 2\Delta\left(\frac{r}{a}\right)^{\alpha}\right) & \text{when } r < a \ (core) \\ n_1(1 - 2\Delta)^{\frac{1}{2}} \approx n_2 & when\ r \geq a \ \text{(cladding)} \end{cases}$$

a = Core radius

n_1= Refractive index of core

n_2 = Refractive index of cladding α = Shape of index profile.

Profile parameter α determines the characteristic refractive index profile of fiber core.

For this reason in this section we consider the waveguiding properties of graded index fiber with a parabolic refractive index profile core. A multimode graded index fiber with a parabolic index profile core is illustrated in Figure 1.16. It may be observed that the meridional rays shown appear to follow curved paths through the fiber core. Using the concepts of geometric optics, the gradual decrease in refractive index from the center of the core creates many refractions of the rays as they are effectively incident on a large number or high to low index interfaces. This mechanism is illustrated in Figure 1.17 where a ray is shown to be gradually curved, with an ever- increasing angle of incidence, until the conditions for total internal reflection are met, and the ray travels back towards the core axis, again being continuously refracted.

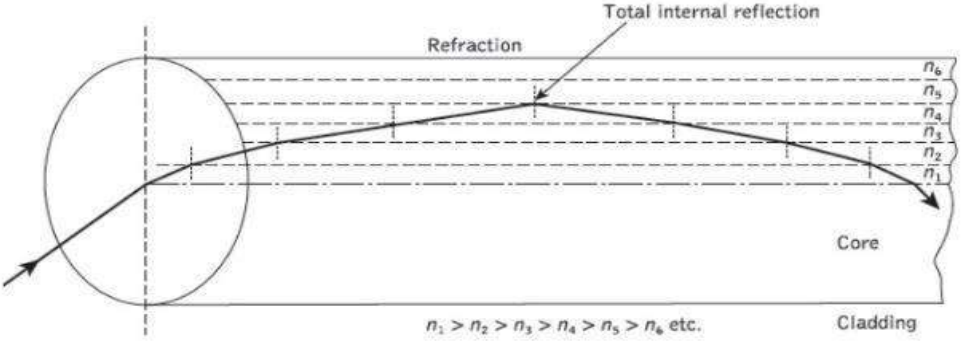


Figure 1.17 An expanded ray diagram showing refraction at the various high to low index interfaces within a graded index fiber, giving an overall curved ray path into the outer regions of the core.

Multimode graded index fibers exhibit far less intermodal dispersion than multimode step index fibers due to their refractive index profile. Although many different modes are excited in the graded index fiber, the different group velocities of the modes tend to be normalized by the index grading. Again considering ray theory, the rays traveling close to the fiber axis have shorter paths when compared with rays which travel

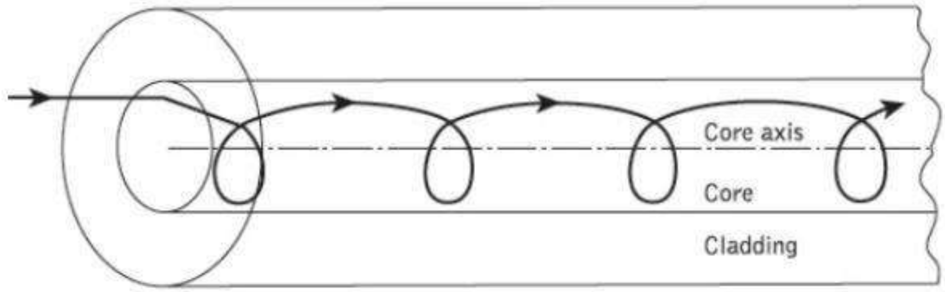


Figure 1.18 A helical skew ray path within a graded index fiber

However, the near axial rays are transmitted through a region of higher refractive index and therefore travel with a lower velocity than the more extreme rays. This compensates for the shorter path lengths and reduces dispersion in the fiber. A similar situation exists for skew rays which follow longer helical paths, as illus- trated in Figure 1.18. These travel for the most part in the lower index region at greater speeds, thus giving the same mechanism of mode transit time equalization. Hence, multi- mode graded index fibers with parabolic or near-parabolic index profile cores have trans- mission bandwidths which may be orders of magnitude greater than multimode step index fiber bandwidths. Consequently, although they are not capable of the bandwidths attain- able with single-mode fibers, such multimode graded index fibers have the advantage of large core diameters (greater than 30 μm) coupled with bandwidths suitable for long- distance communication. The parameters defined for step index fibers (i.e. *NA*, Δ, *V*) may be applied to graded index fibers and give a comparison between the two fiber types. However, it must be noted that for graded index fibers the situation is more complicated since the numerical aperture is a function of the radial distance from the fiber axis. Graded index fibers, therefore, accept less light than corresponding step index fibers with the same relative refractive index difference.

Single-mode fiber

The advantage of the propagation of a single mode within an optical fiber is that the signal dispersion caused by the delay differences between different modes in a multimode fiber may be avoided. Multimode step index fibers do not lend themselves to the propagation of a single mode due to the difficulties of maintaining single-mode operation within the fiber when mode conversion (i.e. coupling) to other guided modes takes place at both input mismatches and fiber imperfections.

Hence, for the transmission of a single mode the fiber must be designed to allow propagation of only one mode, while all other modes are attenuated by leakage or absorption. Following the preceding discussion of multimode fibers, this may be achieved through choice of a suitable normalized frequency for the fiber. For single-mode operation, only the fundamental LP01 mode can exist. Hence the limit of single-mode operation depends on the lower limit of guided propagation for the LP11 mode. The cutoff normalized frequency for the LP11 mode in step index fibers occurs at V_c = 2.405. Thus single-mode propagation of the LP01 mode in step index fibers is possible over the range:

$$0 \leq V < 2.405 \tag{1.51}$$

as there is no cutoff for the fundamental mode. It must be noted that there are in fact two modes with orthogonal polarization over this range, and the term single-mode applies to propagation of light of a particular polarization. Also, it is apparent that the normalized frequency for the fiber may be adjusted to within the range given in Eq. (1.51) by reduction of the core radius.

1. **Cutoff wavelength**

It may be noted that single-mode operation only occurs above a theoretical cutoff wavelength λ_c given by:

$$\lambda_c = \frac{2\pi a n_1}{V_c}(2\Delta)^{\frac{1}{2}} \tag{1.52}$$

where V_c is the cutoff normalized frequency. Hence λ_c is the wavelength above which a particular fiber becomes single-moded.

Dividing Eq. (1.52) by $V = \frac{2\pi}{\lambda} a n_1 (2\Delta)^{\frac{1}{2}}$

$$\frac{\lambda_c}{\lambda} = \frac{V}{V_c} \tag{1.53}$$

Thus for step index fiber where V_c = 2.405, the cutoff wavelength is given by:

$$\lambda_c = \frac{V\lambda}{2.405} \tag{1.54}$$

An effective cutoff wavelength has been defined by the ITU-T which is obtained from a 2 m length of fiber containing a single 14 cm radius loop. This definition was produced because the first higher order LP11 mode is strongly affected by fiber length and curvature near cutoff. Recommended cutoff wavelength values for primary coated fiber range from 1.1 to 1.28 μm for single-mode fiber designed for operation in the 1.3μm wavelength region in order to avoid modal noise and dispersion problems. Moreover, practical transmission systems are generally operated close to the effective cutoff wave-length in order to enhance the fundamental mode confinement, but sufficiently distant from cutoff so that no power is transmitted in the second-order LP11 mode.

2. Mode-field diameter and spot size

Many properties of the fundamental mode are determined by the radial extent of its electromagnetic field including losses at launching and jointing, micro bend losses, waveguide dispersion and the width of the radiation pattern. Therefore, the MFD is an important parameter for characterizing single-mode fiber properties which takes into account the wavelength-dependent field penetration into the fiber cladding. In this context it is a better measure of the functional properties of single-mode fiber than the core diameter. For step index and graded (near parabolic profile) single-mode fibers operating near the cutoff wavelength λc, the field is well approximated by a Gaussian distribution. In this case the MFD is generally taken as the distance between the opposite 1/e = 0.37 field amplitude points and the power 1/e2 = 0.135 points in relation to the corresponding values on the fiber axis.

Another parameter which is directly related to the MFD of a single-mode fiber is the spot size (or mode-field radius) ω0. Hence MFD = 2ω0, where ω0is the nominal half width of the input excitation.

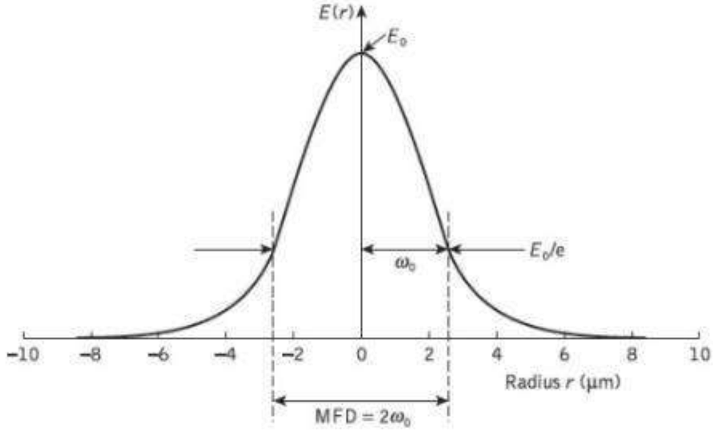


Figure 1.19 Field amplitude distribution $E(r)$ of the fundamental mode in a single-mode fiber illustrating the mode-field diameter (MFD) and spot size (ω_0)

The MFD can therefore be regarded as the single- mode analog of the fiber core diameter in multimode fibers. However, for many refractive index profiles and at typical operating wavelengths the MFD is slightly larger than the single-mode fiber core diameter.

Often, for real fibers and those with arbitrary refractive index profiles, the radial field distribution is not strictly Gaussian and hence alternative techniques have been proposed. However, the problem of defining the MFD and spot size for non-Gaussian field dis- tributions is a difficult one and at least eight definitions exist.

3. Effective refractive index

The rate of change of phase of the fundamental LP01 mode propagating along a straight fiber is determined by the phase propagation constant . It is directly related to the wavelength of the LP01 mode λ01 by the factor 2π, since β gives the increase in phase angle per unit length. Hence:

$$\beta\lambda_{01} = 2\pi \quad \text{or} \quad \lambda_{01} = \frac{2\pi}{\beta} \tag{1.55}$$

Moreover, it is convenient to define an effective refractive index for single-mode fiber, sometimes referred to as a phase index or normalized phase change coefficient n_{eff}, by the ratio of the propagation constant of the fundamental mode to that of the vacuum propagation constant:

$$n_{eff} = \frac{\beta}{k} \tag{1.56}$$

Hence, the wavelength of the fundamental mode λ_{01} is smaller than the vacuum wave-length λ by the factor $1/n_{eff}$ where:

$$\lambda_{01} = \frac{\lambda}{n_{eff}} \tag{1.57}$$

It should be noted that the fundamental mode propagates in a medium with a refractive index *n*(*r*) which is dependent on the distance *r* from the fiber axis. The effective refractive index cantherefore be considered as an average over the refractive index of this medium.

Within a normally clad fiber, not depressed-cladded fibers, at long wavelengths (i.e. small *V* values) the MFD is large compared to the core diameter and hence the electric field extends far into the cladding region. In this case the propagation constant β will be approximately equal to n_2k (i.e. the cladding wave number) and the effective index will be similar to the refractive index of the cladding n_2. Physically, most of the power is transmitted in the cladding material.

At short wavelengths, however, the field is concentrated in the core region and the propagation constant β approximates to the maximum wave number n_1k. Following this discussion, and as indicated previously, then the propagation constant in single-mode fiber varies over the interval $n_2k < \beta < n_1k$. Hence, the effective refractive index will vary over the range $n_2 < n_{eff} < n_1$.

4. Group delay and mode delay factor

The transit time or group delay τ_g for a light pulse propagating along a unit length of fiber is the inverse of the group velocity υ_g . Hence:

$$\tau_g = \frac{1}{\upsilon_g} = \frac{d\beta}{d\omega} = \frac{1}{c}\frac{d\beta}{dk} \qquad (1.61)$$

The group index of a uniform plane wave propagating in a homogeneous medium has been determined as:

$$N_g = \frac{c}{\upsilon_g}$$

However, for a single-mode fiber, it is usual to define an effective group index* N_{ge} By:

$$N_{ge} = \frac{c}{\upsilon_g} \qquad (1.62)$$

Where υ_g is considered to be the group velocity of the fundamental fiber mode. Hence, the specific group delay of the fundamental fiber mode becomes:

$$\tau_g = \frac{N_{ge}}{c} \qquad (1.63)$$

Fiber materials

Most of the fibers are made up of glass consisting of either Silica (SiO_2) or .Silicate. High- loss glass fibers are used for short-transmission distances and low-loss glass fibers are used for long distance applications. Plastic fibers are less used because of their higher attenuation than glass fibers. Glass Fibers.

The glass fibers are made from oxides. The most common oxide is silica whose refractive index is 1.458_at 850 nm. To get different index fibers, the dopants such as GeO_2, P_2O_5 are added to silica. GeO_2 and P_2O_3 increase the refractive index whereas fluorine or B_20_3 decreases the refractive index.

Few fiber compositions are given below as follows,

(i) $GeO_2 – SiO_2$ Core: SiO_2 Cladding

(ii) $P_2Q_5 – SiO_2$, Core; SiO_2 Cladding

The principle raw material for silica is sand. The glass composed of pure silica is referred to as silica glass, nitrous silica or fused silica. Some desirable properties of silica are,

(i) Resistance to deformation at temperature as high as 1000°C.

(ii) High resistance to breakage from thermal shock.

(iii) Good chemical durability.

(iv) High transparency in both the visible and infrared regions.

Basic Requirements and Considerations in Fiber Fabrication

(i) Optical fibers should have maximum reproducibility.

(ii) Fibers should be fabricated with good stable transmission characteristics i.e., the fiber should have invariable transmission characteristics in long lengths.

(iii) Different size, refractive index and refractive index profile, operating wavelengths material. Fiber must be available to meet different system applications.

(iv) The fibers must be flexible to convert into practical cables without any degradation of their characteristics.

(v) Fibers must be fabricated in such a way that a joining (splicing) of the fiber should not affect its transmission characteristics and the fibers may be terminated or connected together with less practical difficulties.

Fiber Fabrication in a Two Stage Process

(i) Initially glass is produced and then converted into perform or rod.

Glass fiber is a mixture of selenides, sulfides and metal oxides. It can be classified into,

1. Halide Glass Fibers
2. Active Glass Fibers
3. Chalgenide Glass Fibers.

Glass is made of pure SiO_2 which refractive index 1.458 at 850 nm. The refractive index of SiO_2 can be increased (or) decreased by adding various oxides are known as dopant. The oxides GeO2 or P2O3 increases the refractive index and B_2O_3 decreases the refractive index of SiO_2 .

The various combinations are,

(i) GeO_2 SiO_2 Core; SiO_2 cladding

(ii) P_2 O_3 – SiO_2 Core; SiO_2 cladding

(iii) SiO_2 Core; B_2O_3, - SiO_2 cladding

(iv) GeO_2- B_2O_3- SiO_2, Core; B_2O_3 - SiO_2 cladding.

From above, the refractive index of core is maximum compared to the cladding.

(1) Halide Glass Fibers

A halide glass fiber contains fluorine, chlorine, bromine and iodine. The most common Halide glass fiber is heavy "metal fluoride glass". It uses ZrF_4 as a major component. This fluoride glass is known by the name ZBLAN Since it is constituents are ZrF_4, BaF_2, LaF_3 $A1F_3$, and NaF.

The percentages of these elements to form ZBLAN fluoride glass is shown as follows,

Materials	Molecular percentage
ZrF_4	54%
BaF_2	20%
LaF_3	4.5%
$A1F_3$	3.5%
NaF	18%

These materials add up to make the core of a glass fiber. By replacing ZrF_4 by HaF_4, the lower refractive index glass is obtained.

The intrinsic losses of these glasses is 0.01 to 0.001 dB/km

(2) Active Glass Fibers

Active glass fibers are formed by adding erbium and neodymium to the glass fibers. The above material performs amplification and attenuation

(3) Chalgenide Glass Fibers

Chalgenide glass fibers are discovered in order to make use of the nonlinear properties of glass fibers. It contains either "S", "Se" or "Te", because they are highly nonlinear and it also contains one element from “Cl”, "Br”, “Cd”,”Ba” or”Si”.The mostly used chalgenide glass is AS_2-S_3, $AS_{40}S_{58}Se_2$ is used to make the core and AS_2S_3 is used to make the cladding material of the glass fiber. The insertion loss is around 1 dB/m.

Plastic Optical Fibers

Plastic optical fibers are the fibers which are made up of plastic material. The core of this fiber is made up of Polymethylmethacrylate (PMMA) or Perflourmated Polymer (PFP).Plastic optical fibers offer more attenuation than glass fiber and is used for short distance applications.

These fibers are tough and durable due to the presence of plastic **material**. The modulus of this plastic material is two orders **of** magnitude lower than that of silica and even a 1 mm diameter graded index plastic optical fiber can be installed **in** conventional fiber cable routes. The diameter of the core of these fibers is 10-20 times larger than that of glass fiber which reduces the connector losses without sacrificing coupling efficiencies. So we can use inexpensive connectors, splices and transceivers made up of plastic injection-molding technology. Graded index plastic optical fiber is in great demand in customer premises to deliver high-speed services due to its high bandwidth.

UNIT-II

SIGNAL DISTORTION IN OPTICAL FIBERS

Introduction

One of the important property of optical fiber is signal attenuation. It is also known as fiber loss or signal loss. The signal attenuation of fiber determines the maximum distance between transmitter and receiver. The attenuation also determines the number of repeaters required, maintaining repeater is a costly affair. Another important property of optical fiber is distortion mechanism. As the signal pulse travels along the fiber length it becomes more broader. After sufficient length the broad pulses starts overlapping with adjacent pulses. This creates error in the receiver. Hence the distortion limits the information carrying capacity of fiber.

Attenuation

- Attenuation is a measure of decay of signal strength or loss of light power that occurs as light pulses propagate through the length of the fiber.
- In optical fibers the attenuation is mainly caused by two physical factors absorption and scattering losses. Absorption is because of fiber material and scattering due to structural imperfection within the fiber. Nearly 90 % of total attenuation is caused by Rayleigh scattering only. Microbending of optical fiber also contributes to the attenuation of signal.
- The rate at which light is absorbed is dependent on the wavelength of the light and the characteristics of particular glass. Glass is a silicon compound, by adding different additional chemicals to the basic silicon dioxide the optical properties of the glass can be changed.
- The Rayleigh scattering is wavelength dependent and reduces rapidly as the wavelength of the incident radiation increases.

- The attenuation of fiber is governed by the materials from which it is fabricated, the manufacturing process and the refractive index profile chosen. Attenuation loss is measured in dB/km.

Attenuation Units

As attenuation leads to a loss of power along the fiber, the output power is significantly less than the couples power. Let the couples optical power is p(0) i.e. at origin (z = 0).

Then the power at distance z is given by,

$$P(z) = P(0)e^{-\alpha_p z} \quad \dots (2.1.1)$$

where, α_p is fiber attenuation constant (per km).

$$\alpha_p = \frac{1}{z} \ln \left[\frac{P(0)}{P(z)}\right]$$

$$\boxed{\alpha_{dB/km} = 10 . \frac{1}{z} \log \left[\frac{P(0)}{P(z)}\right]}$$

$$\alpha_{dB/km} = 4.343 \ \alpha_p \ \text{per km}$$

This parameter is known as fiber loss or fiber attenuation.

- Attenuation is also a function of wavelength. Optical fiber wavelength as a function of wavelength is shown in Fig. 2.1.1.

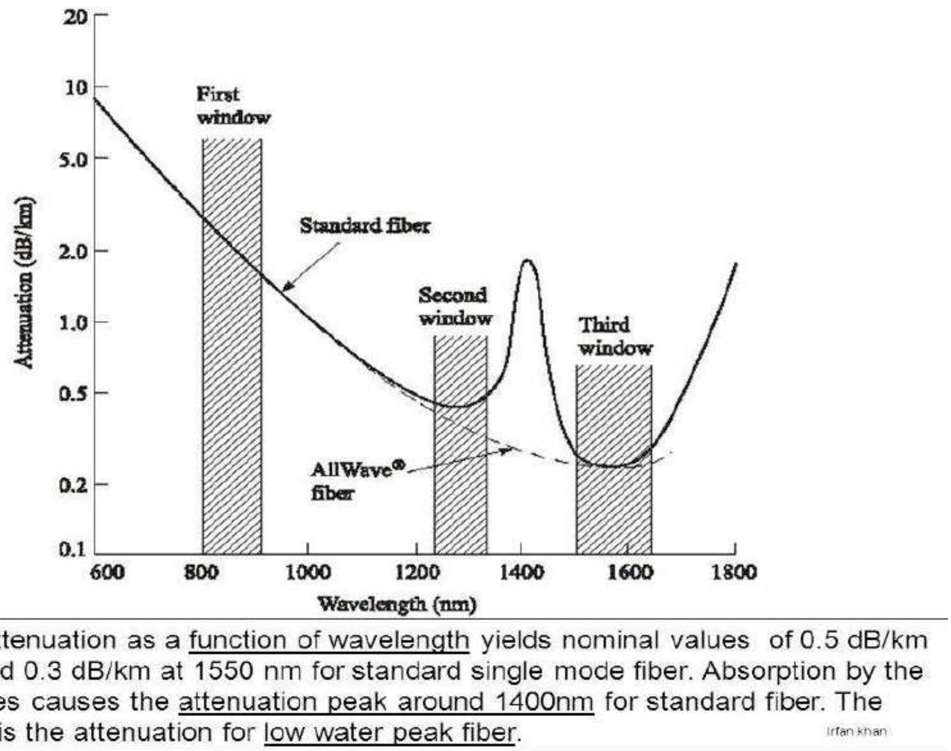


Fig 2.1.1: Optical fiber wavelength as a function of wavelength

Example 2.1.1 : A low loss fiber has average loss of 3 dB/km at 900 nm. Compute the length over which –

a) Power decreases by 50 %   b) Power decreases by 75 %.

Solution : α = 3 dB/km

$\Rightarrow \quad \frac{p(0)}{p(z)} = 50\ \% = 0.5$

a) Power decreases by 50 %.

□ is given by,

$$\left[\frac{200\ \mu W}{P(z)}\right] = 10^{2.4}$$

$$3 = 10 \cdot \frac{1}{z} \log [0.5]$$

z = **1 km... Ans.**

b) $\frac{P(0)}{P(z)} = 25\ \% = 0.25$ Since power decrease by 75

%.

$$3 = 10 \text{ x } \frac{1}{z} \log[0.25]$$

z = **2 km... Ans.**

Example 2.1.2 : For a 30 km long fiber attenuation 0.8 dB/km at 1300nm. If a 200 µwatt power is launched into the fiber, find the output power.

Solution : z = 30 km

 = 0.8 dB/km

P(0) = 200

µW

Attenuation in optical fiber is given by,

$$0.8 = 10 \text{ x } \frac{1}{30} \log \left[\frac{200\ \mu W}{P(z)}\right]$$

$$2.4 = 10 \text{ x } \log \left[\frac{200\ \mu W}{P(z)}\right]$$

Example 2.1.3 : When mean optical power launched into an 8 km length of fiber is 12 µW, the mean optical power at the fiber output is 3 µW.

Determine –

Overall signal attenuation in dB.

The overall signal attenuation for a 10 km optical link using the same fiber with splices at 1 km intervals, each giving an attenuation of 1 dB.

Solution: Given : z = 8 km

P(0) = 120 µW

P(z) = 3 µW

1) Overall attenuation is given by,

$$\alpha = 10\ .\log\left[\frac{P(0)}{P(z)}\right]$$

$$\alpha = 10\ .\log\left[\frac{120}{3}\right]$$

$$\alpha = 16.02\ \text{dB}$$

2) Overall attenuation for 10 km,

Attenuation per km

$\propto_{dB} = \frac{16.02}{z} = \frac{16.02}{8} = 2.00\ \text{dB/km}$ Attenuation in 10 km link = 2.00 x 10 = 20 dB

In 10 km link there will be 9 splices at 1 km interval. Each splice introducing attenuation of 1 dB.

Total attenuation = 20 dB + 9 dB = **29 dB**

Example 2.1.4 : A continuous 12 km long optical fiber link has a loss of 1.5 dB/km.

- What is the minimum optical power level that must be launched into the fiber to maintain as optical power level of 0.3 μW at the receiving end?
- What is the required input power if the fiber has a loss of 2.5 dB/km?

Solution : Given data : z = 12 km

= 1.5 dB/km

P(0) = 0.3 μW

- Attenuation in optical fiber is given by,

$$\alpha = 10 \text{ x } \frac{1}{z} \log\left(\frac{P(0)}{P(z)}\right)$$

$$1.5 = 10 \text{ x } \frac{1}{12} \log\left(\frac{0.3\ \mu W}{P(z)}\right)$$

$$\log\left(\frac{0.3\ \mu W}{P(z)}\right) = \frac{1.5}{0.833}$$

= 1.80

$$\left(\frac{0.3\ \mu W}{P(z)}\right) = 10^{1.8}$$

$$P(z) = \left(\frac{0.3\ \mu W}{10^{1.8}}\right) = \frac{0.3}{63.0}$$

$$P(z) = 4.76 \times 10^{-9} W$$

Optical power output = **4.76 x 10^{-9} W** **… Ans.**

ii) Input power = ? P(0)

When α = 2.5 dB/km

$$\alpha = 10 \times \frac{1}{z} \log\left(\frac{P(0)}{P(z)}\right)$$

$$2.5 = 10 \times \frac{1}{z} \log\left(\frac{P(0)}{4.76 \times 10^{-9}}\right)$$

$$\log\left(\frac{P(0)}{4.76 \times 10^{-9}}\right) = \frac{2.5}{0.833} = 3$$

$$\frac{P(0)}{4.76 \times 10^{-9}} = 10^3 = 1000$$

P(0) = 4.76 µW

Input power= **4.76 µW** **… Ans.**

Example 2.1.5 : Optical power launched into fiber at transmitter end is 150 µW. The power at the end of 10 km length of the link working in first windows is – 38.2 dBm. Another system of same length working in second window is 47.5 µW. Same length system working in third window has 50 % launched power. Calculate fiber attenuation for each case and mention wavelength of operation. **[Jan./Feb.-2009, 4 Marks]**

Solution : Given data:

$P(0) = 150\ \mu W$

$z = 10\ km$

$$P(z) = -38.2\ dBm \Rightarrow \begin{cases} -38.2 = 10 \log \dfrac{P(z)}{1\ mW} \\ P(z) = 0.151\ \mu W \end{cases}$$

$z = 10\ km$

$$\boxed{\alpha = 10 \times \frac{1}{z} \log \left[\frac{P(0)}{P(z)}\right]}$$

Attenuation in 1st window:

$$\alpha_1 = 10 \times \frac{1}{10} \log \left[\frac{150}{0.151}\right]$$

$$\alpha_1 = 2.99 \text{ dB/km}$$

Attenuation in 2^{nd} window:

$$\alpha_2 = 10 \times \frac{1}{10} \log \left[\frac{150}{47.5}\right]$$

$$\alpha_2 = 0.49 \text{ dB/km}$$

Attenuation in 3^{rd} window:

$$\alpha_3 = 10 \times \frac{1}{10} \log \left[\frac{150}{75}\right]$$

$$\alpha_3 = 0.30 \text{ dB/km}$$

Wavelength in 1^{st} window is 850 nm.

Wavelength in 2^{nd} window is 1300 nm.

Wavelength in 3^{rd} window is 1550 nm.

Example 2.1.6 : The input power to an optical fiber is 2 mW while the power measured at the output end is 2 µW. If the fiber attenuation is 0.5 dB/km, calculate the length of the fiber.

Solution: Given : P(0) = 2 mwatt = 2×10^{-3} watt

P(z) = 2 µwatt = 2×10^{-6} watt

α = 0.5 dB/km

structure of fiber. The damages are proportional to the intensity of ionizing particles. This results in increasing attenuation due to atomic defects and absorbing optical energy. The total dose a material receives is expressed in rad (Si), this is the unit for measuring radiation absorbed in bulk silicon.

1 rad (Si) = 0.01 J.kg

The higher the radiation intensity more the attenuation as shown in Fig 2.2.1.

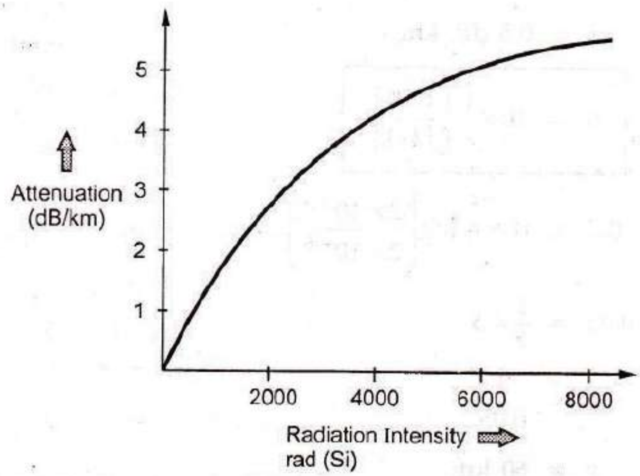


Fig. 2.2.1 Ionizing radiation intensity Vs fiber attenuation

Extrinsic Absorption

Extrinsic absorption occurs due to electronic transitions between the energy level and because of charge transitions from one ion to another. A major source of attenuation is from transition of metal impurity ions such as iron, chromium, cobalt and copper. These losses can be upto 1 to 10 dB/km. The effect of metallic impurities can be reduced by glass refining techniques.

- Another major extrinsic loss is caused by absorption due to **OH (Hydroxil)** ions impurities dissolved in glass. Vibrations occur at wavelengths between 2.7 and 4.2 μm.

  The absorption peaks occurs at 1400, 950 and 750 nm. These are first, second and third overtones respectively.

- Fig. 2.2.2 shows absorption spectrum for OH group in silica. Between these absorption peaks there are regions of low attenuation.

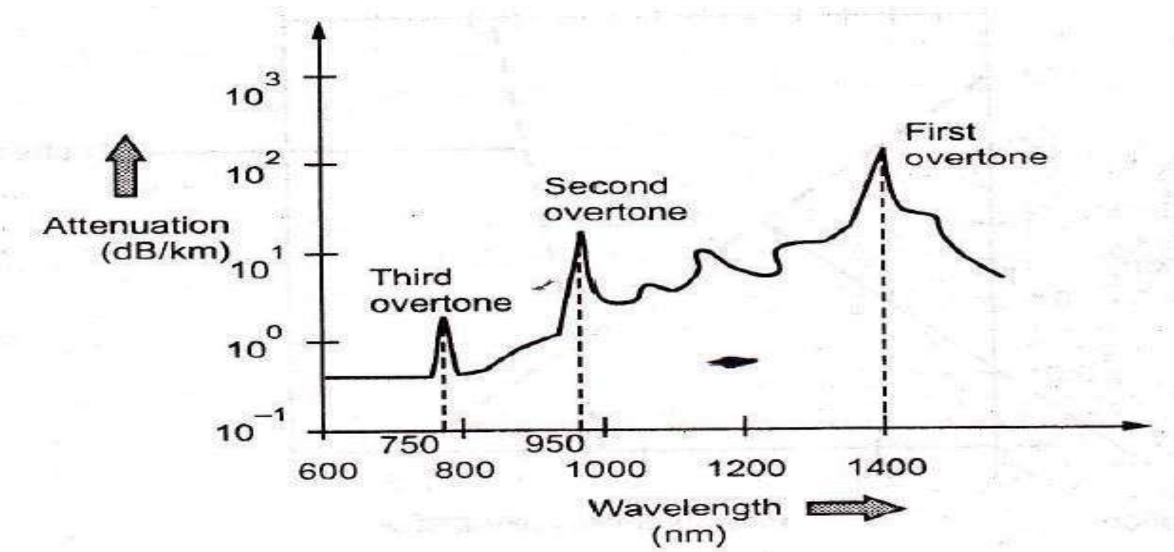


Fig. 2.2.2 Absorption spectra for OH group

Intrinsic Absorption

Intrinsic absorption occurs when material is in absolutely pure state, no density variation and inhomogenities. Thus intrinsic absorption sets the fundamental lower limit on absorption for any particular material.

- Intrinsic absorption results from electronic absorption bands in UV region and from atomic vibration bands in the near infrared region.
- The electronic absorption bands are associated with the band gaps of amorphous glass materials. Absorption occurs when a photon interacts with an electron in the valene band and excites it to a higher energy level. UV absorption decays exponentially with increasing wavelength (λ).

- In the IR (infrared) region above 1.2 μm the optical waveguide loss is determined by presence of the OH ions and inherent IR absorption of the constituent materials. The inherent IR absorption is due to interaction between the vibrating band and the electromagnetic field of optical signal this results in transfer of energy from field to the band, thereby giving rise to absorption, this absorption is strong because of many bonds present in the fiber.

The ultraviolet loss at any wavelength is expressed as,

$$\propto_{uv} = \frac{154.2}{46.6\,x + 60} \times 10^{-2} \times e^{\left(\frac{4.63}{\lambda}\right)} \qquad \text{... (2.2.1)}$$

where, x is mole fraction of GeO_2.

λ is operating wavelength.

α_{uv} is in dB/km.

The loss in infrared (IR) region (above 1.2 μm) is given by expression :

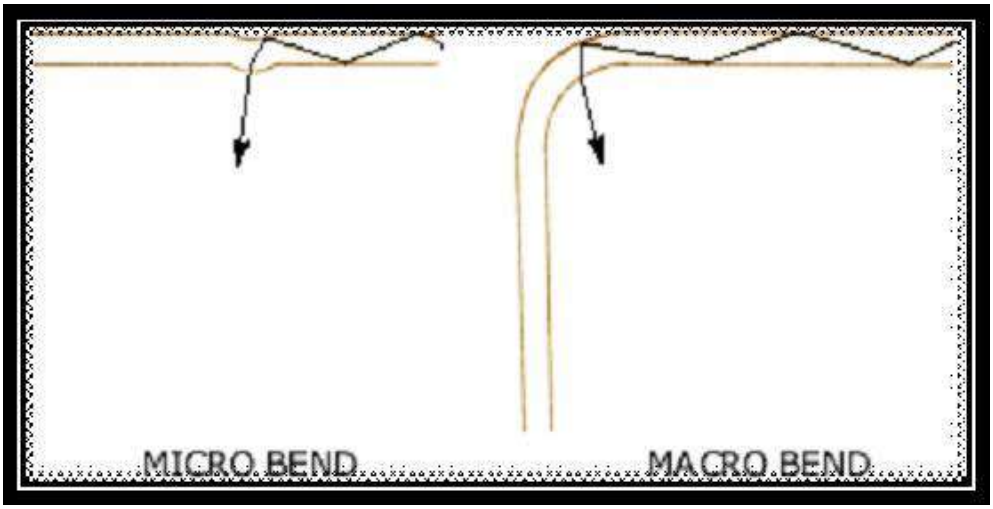
$$\propto_{IR} = 7.81 \times 10^{11} \times e^{\left(\frac{-48.48}{\lambda}\right)} \qquad \text{... (2.2.2)}$$

The expression is derived for GeO_2-SiO_2 glass fiber.

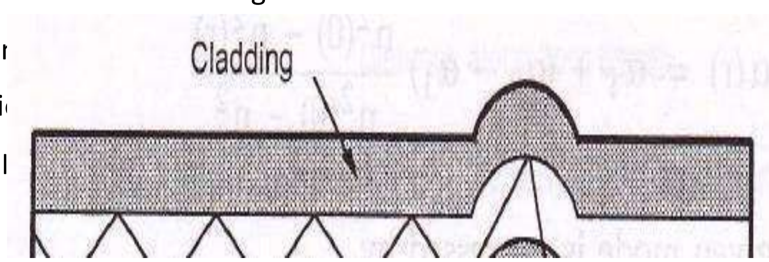
Rayleigh Scattering Losses

Scattering losses exists in optical fibers because of microscopic variations in the material density and composition. As glass is composed by randomly connected network of molecules and several oxides (e.g. SiO_2, GeO_2 and P_2O_5), these are the major cause of compositional structure fluctuation. These two effects results to variation in refractive index and Rayleigh type scattering of light.

Micro bending losses can be minimized by placing a compressible jacket over the fiber. When external forces are applied to this configuration, the jacket will be deformed but the fiber will tend to stay relatively straight.



- Microbending is a loss due to small bending or distortions. This small microbending is not visible. The losses due to this are temperature related, tensile related or crush related.
- The effects of microbending on multimode fiber can result in increasing attenuation (dependi ghs on the spectral attenuati llation and testing. Fig.2.4.2 il



Macrobending

For slight bends, the loss is extremely small and is not observed. As the radius of curvature decreases, the loss increases exponentially until at a certain critical radius of curvature loss becomes observable. If the bend radius is made a bit smaller once this threshold point has been reached, the losses suddenly become extremely large. It is known that any bound core mode has an evanescent field tail in the cladding which decays exponentially as a function of distance from the core. Since this field tail moves along with the field in the core, part of the energy of a propagating mode travels in the fiber cladding. When a fiber is bent, the field tail on the far side of the centre of curvature must move faster to keep up with the field in the core, for the lowest order fiber mode. At a certain critical distance $x_{c,}$ from the centre of the fiber; the field tail would have to move faster than the speed of light to keep up with the core field. Since this is not possible the optical energy in the field tail beyond x_c radiates away.
The amount of optical radiation from a bent fiber depends on the field strength at x_c and on the radius of curvature R. Since higher order modes are bound less tightly to the fiber core than lower order modes, the higher order modes will radiate out of the fiber first.

- The change in spectral attenuation caused by macrobending is different to microbending. Usually there are no peaks and troughs because in a macrobending no light is coupled back into the core from the cladding as can happen in the case of microbends.

- The macrobending losses are cause by large scale bending of fiber. The losses are eliminated when the bends are straightened. The losses can be minimized by not exceeding the long term bend radii. Fig. 2.4.3 illustrates macrobending.

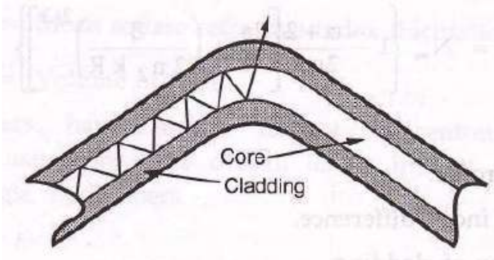
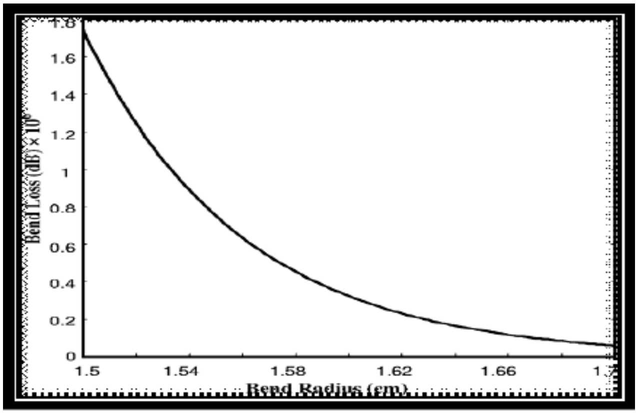


Fig. 2.4.3 Macrobending



Macro bending Loss

Core and Cladding Loss

- Since the core and cladding have different indices of refraction hence they have different attenuation coefficients α_1 and α_2 respectively.

$$\propto(r) = \propto_1 + (\propto_2 - \propto_1)\frac{n^2(0) - n^2(r)}{n^2(0) - n_2^2}$$

- For step index fiber, the loss for a mode order (v, m) is given by,

$$\propto_{vm} = \propto_1 \frac{P_{core}}{P} + \propto_2 \frac{P_{cladding}}{P}$$ ...

(2.5.1)

For low-order modes, the expression reduced to

$$\propto_{vm} = \propto_1 + (\propto_2 + \propto_1)\frac{P_{cladding}}{P}$$ ...

(2.5.2)

where, $\frac{P_{core}}{P}$ and $\frac{P_{cladding}}{P}$ are fractional powers.

- For graded index fiber, loss at radial distance is expressed as,

... (2.5.3)

The loss for a given mode is expressed by,

$$\propto_{Graded\ Index} = \frac{\int_0^\infty \propto(r)\, P(r)\, r\, dr}{\int_0^\infty P(r)\, r\, dr}$$ ...

(2.5.4)

where, P(r) is power density of that model at radial distance r.

Signal Distortion in Optical Waveguide

- The pulse gets distorted as it travels along the fiber lengths. Pulse spreading in fiber is referred as dispersion. Dispersion is caused by difference in the propagation times of light rays that takes different paths during the propagation. The light pulses travelling down the fiber encounter dispersion effect because of this the pulse spreads out in time domain. Dispersion limits the information bandwidth. The distortion effects can be analyzed by studying the group velocities in guided modes.

Information Capacity Determination

- Dispersion and attenuation of pulse travelling along the fiber is shown in Fig. 2.6.1.

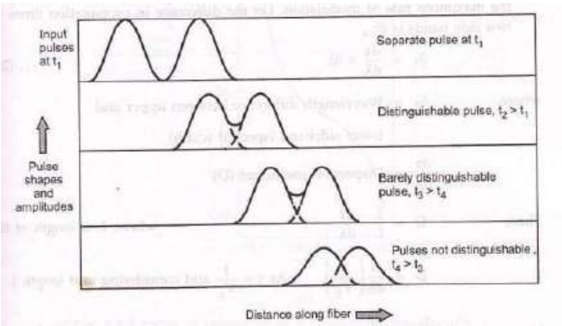


Fig. 2.6.1 Dispersion and attenuation in fiber

- Fig. 2.6.1 shows, after travelling some distance, pulse starts broadening and overlap with the neighbouring pulses. At certain distance the pulses are not even distinguishable and error will occur at receiver. Therefore the information capacity is specified by bandwidth- distance product (MHz . km). For step index bandwidth distance product is 20 MHz . km and for graded index it is 2.5 MHz . km.

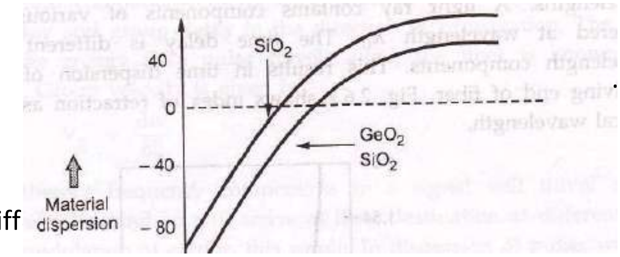
Group Delay

- Consider a fiber cable carrying optical signal equally with various modes and each mode contains all the spectral components in the wavelength band. All the spectral components travel independently and they observe different **time delay** and **group delay** in the direction of propagation. The velocity at which the energy in a pulse travels along the fiber is known as **group velocity**. Group velocity is given by,

$$V_g = \frac{\partial w}{\partial \beta} \qquad \ldots (2.6.1)$$

- Thus different frequency components in a signal will travel at different group velocities and so will arrive at their destination at different times, for digital modulation of carrier, this results in dispersion of pulse, which affects the maximum rate of modulation. Let the difference in propagation times for two side bands is $\delta\tau$.

$$\delta\tau = \frac{d\tau}{d\lambda} \, x \, \delta\lambda$$



…

(2.6.2)

where, $\delta\tau$ = Wavelength diff (spectral width) $\frac{d\tau}{d\lambda}$ = Dispersion coefficient (D)

Then, $D = \frac{1}{L} . \frac{d\tau}{d\lambda}$ where, L is length of fiber.

$$D = \frac{d}{d\lambda}\left(\frac{1}{V_g}\right)$$

As $\tau = \frac{1}{V_g}$ and considering unit length L = 1.

Now $\frac{1}{V_g} = \frac{d\beta}{d\omega}$

$$\frac{1}{V_g} = \frac{d\lambda}{d\omega} x \frac{d\beta}{d\lambda}$$

$$\frac{1}{V_g} = \frac{-\lambda^2}{2\pi c} x \frac{d\beta}{d\lambda}$$

$$D = \frac{d}{d\lambda}\left(\frac{-\lambda^2}{2\pi c} . \frac{d\beta}{d\lambda}\right)$$

- Dispersion is measured in picoseconds per nanometer per kilometer.

Material Dispersion

Material dispersion is also called as chromatic dispersion. Material dispersion exists due to change in index of refraction for different wavelengths. A light ray contains components of various wavelengths centered at wavelength λ10. The time delay is different for different wavelength components. This results in time dispersion of pulse at the receiving end of fiber. Fig. 2.6.2 shows index of refraction as a function of optical wavelength.

The material dispersion for unit length (L = 1) is given by

$$D_{mat} = \frac{-\lambda}{c} x \frac{d^2n}{d\lambda^2}$$

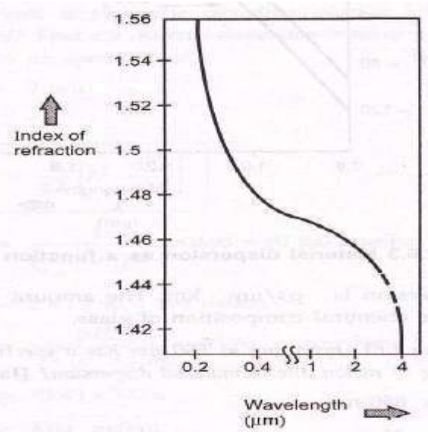


Fig. 2.6.2 Index of refraction as a function of wavelength

... (2.6.4)

where, c = Light velocity

λ = Center wavelength

$\frac{d^2n}{d\lambda^2}$ = Second derivative of index of refraction w.r.t wavelength

Negative sign shows that the upper sideband signal (lowest wavelength) arrives before the lower sideband (highest wavelength).

- The unit of dispersion is : ps/nm . km. The amount of material dispersion depends upon the chemical composition of glass.

Example 2.6.1 : An LED operating at 850 nm has a spectral width of 45 nm. What is the pulse spreading in ns/km due to material dispersion?

Solution : Given : λ = 850 nm

σ = 45 nm

pulse broadening due to material dispersion is given by,

$\sigma_m = \sigma LM$

Considering length L = 1 metre

$$\text{Material dispersion constant } D_{mat} = \frac{-\lambda}{c} \cdot \frac{d^2n}{d\lambda^2}$$

For LED source operating at 850 nm, $\left|\lambda^2 \frac{d^2n}{d\lambda^2}\right|$ =0.025

$$M = \frac{1}{c\lambda}\left|\lambda^2 \frac{d^2n}{d\lambda^2}\right| = \frac{1}{(3 \times 10^5)(850)} \times 0.025$$

M = 9.8 ps/nm/km

σ_m = **441 ns/km** ...

Ans.

Waveguide Dispersion

- Waveguide dispersion is caused by the difference in the index of refraction between the core and cladding, resulting in a 'drag' effect between the core and cladding portions of the power.
- Waveguide dispersion is significant only in fibers carrying fewer than 5-10 modes. Since multimode optical fibers carry hundreds of modes, they will not have observable waveguide dispersion.
- The group delay (τ_{wg}) arising due to waveguide dispersion.

$$(\tau_{wg}) = \frac{L}{c}\left[n_2 + n_2 \Delta \frac{d\,(kb)}{dk}\right]$$ ...

(2.6.5)

Where, b = Normalized propagation constant k = 2π / λ (group velocity)

Normalized frequency V,

$$V = ka(n_1^2 - n_2^2)^{\frac{1}{2}}$$

$$\tau_{wg} = \frac{L}{c}\left[n_2 + n_2 \Delta \frac{d\,(V_b)}{dV}\right]$$

The second term $\frac{d(V_b)}{dV}$ is waveguide dispersion and is mode dependent term..

- As frequency is a function of wavelength, the group velocity of the energy varies with frequency. The produces additional losses (waveguide dispersion). The propagation constant (b) varies with wavelength, the causes of which are independent of material dispersion.

Chromatic Dispersion

- The combination of material dispersion and waveguide dispersion is called chromatic dispersion. These losses primarily concern the spectral width of transmitter and choice of correct wavelength.

- A graph of effective refractive index against wavelength illustrates the effects of material, chromatic and waveguide dispersion.

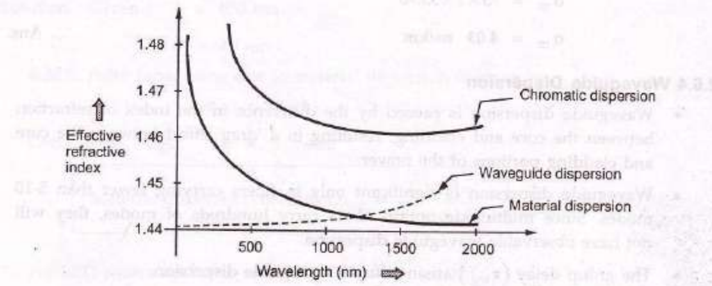


Fig. 2.6.4 Graph of effective refractive index against wavelength showing effects of chromatic, waveguide and material dispersion

Material dispersion and waveguide dispersion effects vary in vary in opposite senses as the wavelength increased, but at an optimum wavelength around 1300 nm, two effects almost cancel each other and chromatic dispersion is at minimum. Attenuation is therefore also at minimum and makes 1300 nm a highly attractive operating wavelength.

Modal Dispersion

As only a certain number of modes can propagate down the fiber, each of these modes carries the modulation signal and each one is incident on the boundary at a different angle, they will each have their own individual propagation times. The net effect is spreading of pulse, this form o dispersion is called modal dispersion.

- Modal dispersion takes place in multimode fibers. It is moderately present in graded index fibers and almost eliminated in single mode step index fibers.

- Modal dispersion is given by,

$$\Delta t_{modal} = \frac{n_1 Z}{c}\left(\frac{\Delta}{1-\Delta}\right)$$

where Δt_{modal} = Dispersion

n_1 = Core refractive index

Z = Total fiber length

c = Velocity of light in air

□= Fractional refractive index $\left(\frac{n_1 - n_2}{n_1}\right)$

Putting in above equation $\Delta = \frac{(NA^2)Z}{2n_1 c}$

$$\Delta t_{modal} = \frac{(NA^2)Z}{2n_1\ c}$$

$$t_{r\ mod} = 0.44\ (\Delta t_{modal})\pi r^2$$

Example 2.6.3 : For a single mode fiber n_2 = 1.48 and = 0.2 % operating at A = 1320 nm, compute the waveguide dispersion if $V.\frac{d^2(Vb)}{dv^2} = 0.26.$

Solution : n_2 = 1.48

0.2

= 1320 nm

Waveguide dispersion is given by,z

$$D_{wg}(\lambda) = \frac{-n_2\Delta}{c\lambda}\left[V\frac{d^2(Vb)}{dV^2}\right]$$

$$= \frac{-1.48 \times 0.2}{3 \times 10^5 \times 1320}[0.20]$$

-1.943 picosec/nm .

km.

Higher Order Dispersion

S.

$$S = \frac{dD}{d\lambda}$$

Higher order dispersive effective effects are governed by dispersion slope

where, D is total dispersion

Also,

$S = \left(\frac{2\pi c}{\lambda^2}\right)^2 \beta_3 + \left(\frac{4\pi c}{\lambda^3}\right)\beta_2$ where,

β_2 and β_3 are second and third order dispersion parameters.

- Dispersion slope S plays an important role in designing WDM system

Dispersion Induced Limitations

- The extent of pulse broadening depends on the width and the shape ofinput pulses. The pulse broadening is studied with the help of wave equation.

Basic Propagation Equation

- The basic propagation equation which governs pulse evolution in a single mode fiber is given by,

$$\frac{\partial A}{\partial z} + \beta_1 \frac{\partial A}{\partial t} + \frac{i\beta_2}{2} . \frac{\partial^2 A}{\partial t^2} - \frac{\beta_3}{6} \frac{\partial^3 A}{\partial t^3} = 0$$

where,

β_1, β_2 and β_3 are different dispersion parameters.

Chirped Gaussian Pulses

- A pulse is said to b e chirped if its carrier frequency changes with time.
- For a Gaussian spectrum having spectral width σ_ω, the pulse broadening factor is given by,

$$\frac{\sigma^2}{\sigma_0^2} = \left(1 + \frac{C\beta_2 L}{2\sigma_2^2}\right)^2 + (1 + V_\omega^2)\left(\frac{\beta_2 L}{2\sigma_0^2}\right)^2 + (1 + C + V_\omega^2)^2 \left(\frac{\beta_3 L}{4\sqrt{2\sigma_0^3}}\right)\pi r^2$$

where, $V_\omega = 2\sigma_\omega \sigma_0$

Limitations of Bit Rate

- The limiting bit rate is given by,

$$4B\sigma \leq 1$$

- The condition relating bit rate-distance product (BL) and dispersion (D) is given

$$BL\ |S|\sigma_\lambda^2 \leq \frac{1}{\sqrt{8}}$$

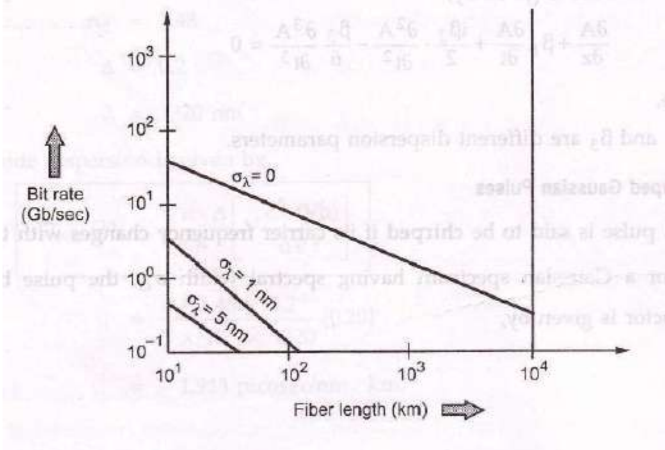


Fig. 2.6.5 Dependence of bit rate on fiber length

where, S is dispersion slope.Limiting bit rate a single mode fibers as a function of fiber length for σλ = 0, a and 5nm is shown in fig. 2.6.5.

Polarization Mode Dispersion (PMD)

- Different frequency component of a pulse acquires different polarization state (such as linear polarization and circular polarization). This results in pulse broadening is kno**w as polarization mode dispersion (PMD).**
- PMD is the limiting factor for optical communication system at high data rates. The effects of PMD must be compensated.

Pulse Broadening in GI Fibers

- The core refractive index varies radially in case of graded index fibers, hence it supports multimode propagation with a low intermodal delay distortion and high data rate over long distance is possible. The higher order modes travelling in outer regions of the core, will travel faster than the lower order modes travelling in high refractive index region. If the index profile is carefully controlled, then the transit times of the individual modes will be identical, so eliminating modal dispersion.
    - The r.m.s pulse broadening is given as:

$$\sigma = \left(\sigma^2_{intermodal} + \sigma^2_{intermodal}\right)^{1/2} \quad \text{... (2.7.1)}$$

where,

$\sigma_{intermodal}$ – R.M.S pulse width due to intermodal delay distortion.

$\sigma_{intermodal}$ – R.M.S pulse width resulting from pulse broadening within each mode.

- The intermodal delay and pulse broadening are related by expression given by Personick.

$$\sigma_{intermodal} = \left(\langle\tau_g^2\rangle - \langle\tau_g\rangle^2\right)^{1/2} \quad \text{... (2.7.2)}$$

Where τ_g is group delay.

From this the expression for intermodal pulse broadening is given as:

$$\propto_{intermodal} = \frac{LN_1\Delta}{2c} \cdot \frac{\propto}{\propto + 1}\left(\frac{\propto + 2}{3\propto + 2}\right)^{1/2} x$$

$$\left[c_1^2 + \frac{4c_1c_2(\propto+1)}{2\propto+1} + \frac{16\Delta^2c_2^2(\propto+1)^2}{(5\propto+2)(3\propto+2)}\right]^{1/2}$$

... (2.7.3)

$$c_1 = \frac{\propto - 2 - E}{\propto + 2} \; and \; c_2 = \frac{3\propto - 2 - 2c}{2(\propto + 2)}$$

- The intramodal pulse broadening is given as :

$$\sigma^2_{intramodal} = \left(\frac{\sigma\lambda}{\lambda}\right)^2 \left\langle\left(\lambda\frac{d\tau g}{d\lambda}\right)^2\right\rangle$$

... (2.7.4)

Where σλ is spectral width of optical

source.

Solving the expression

gives :

$$\sigma^2_{intramodal} = \frac{L}{c} \cdot \frac{\sigma\lambda}{\lambda}\left[\left(-\lambda^2\frac{d^2n_1}{d\lambda^2}\right)^2 - N_1c_1\Delta\right.$$

$$\left(2\lambda^2 \frac{d^2 n_1}{d\lambda^2} \cdot \frac{\alpha}{\alpha+1} - N_1 c_1 \Delta \frac{4\alpha^2}{(\alpha+2)(3\alpha+2)}\right)\Bigg]^{1/2}$$

Mode Coupling

After certain initial length, the pulse distortion increases less rapidly because of mode coupling. The energy from one mode is coupled to other modes because of Structural imperfections, Fiber diameter variations, Refractive index variations, Microbends in cable.Due to the mode coupling, average propagation delay become less and intermodal distortion reduces.Suppose certain initial coupling length = L_C, mode coupling length, over L_C = Z. Additional loss associated with mode coupling = h (dB/ km).Therefore the excess attenuation resulting from mode coupling = hZ. The improvement in pulse spreading by mode coupling is given as :

$$hZ\left(\frac{\sigma_c}{\sigma_0}\right) = C$$

where, C is constant independent of all dimensional quantities and refractive indices. σ_C is pulse broadening under mode coupling.σ_0 is pulse broadening in absence of mode coupling. For long fiber length's the effect of mode coupling on pulse distortion is significant. For a graded index fiber, the effect of distance on pulse broading for various coupling losses are shown

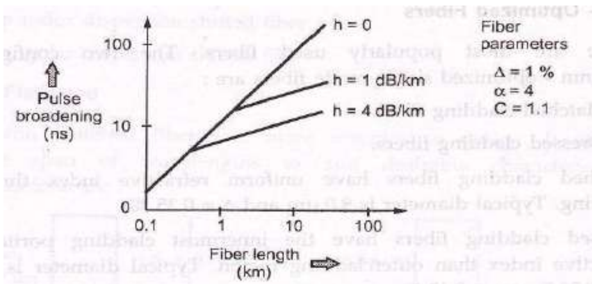


Fig. 2.8.1 Mode coupling effects on pulse broadening

Design Optimization

Features of single mode fibers are : Longer life, Low attenuation ,Signal Transfer quality is good, Modal noise is absent,Largest BW-distance product.

Basic design – optimization includes the following : Dispersion,Mode field,Diameter,bending loss, Refractive index profile,Cut-off wavelength.

Refractive Index Profile

Dispersion of single mode silica fiber is lowest at 1300 nm while its attenuation is minimum at 1550 nm. For archiving maximum transmission distance the dispersion null should be at the wavelength of minimum attenuation. The waveguide dispersion is easier to control than the material dispersion. Therefore a variety of core-cladding refractive.

1300 nm – Optimized Fibers

These are most popularly used fibers. The two configurations of 1300 nm – optimized single mode fibers are

- Matched cladding fibers.
- Dressed cladding fibers.

Matched cladding fibers have uniform refractive index throughout its cladding. Typical diameter is 9.0 μm and Δ = 0.35 %.

Dressed cladding fibers have the innermost cladding portion has low refractive index than outrcladding region. Typical diameter is 8.4 μm and Δ_1 = 0.25 %, Δ_2 = 0.12 %.

Fig 2.9.1 shows both types of fibers.

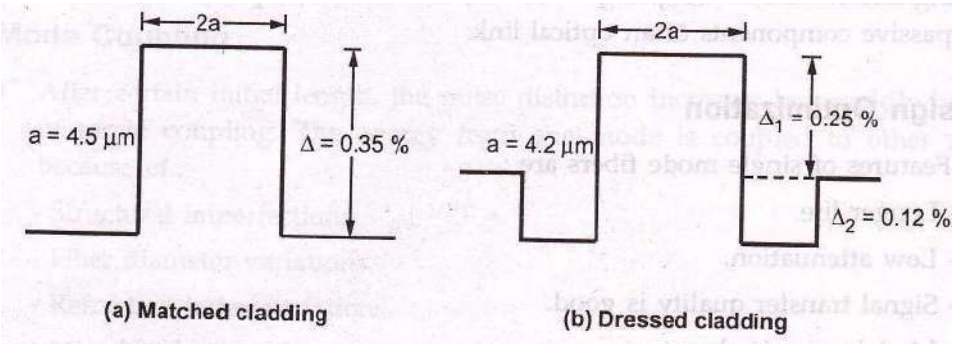


Fig. 2.9.1 1300 nm - optimized refractive index profile

Dispersion Shifted Fibers

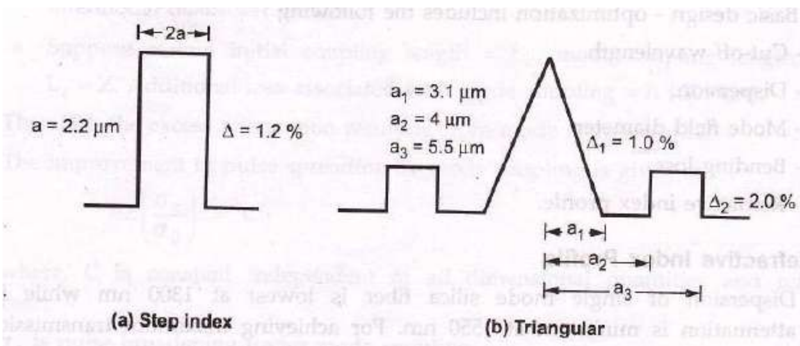
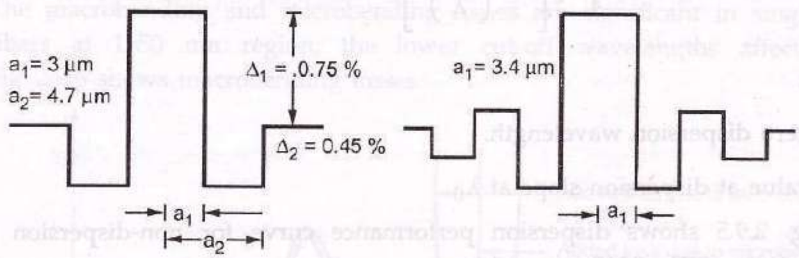


Fig. 2.9.2 Dispersion shifted fibers

2. The addition of wavelength and material dispersion can shift the zero dispersion point of longer wavelength. Two configurations of dispersion shifted fibers are

Dispersion Flattened

Dispersion flattened fibers are more complex to design. It offers much broader span of wavelengths to suit desirable characteristics. Two configurations are :



(a) Double clad profile (b) Quadruple clad profile

Fig. 2.9.3 Dispersion flattened

- Fig 2.9.4 shows total resultant dispersion.

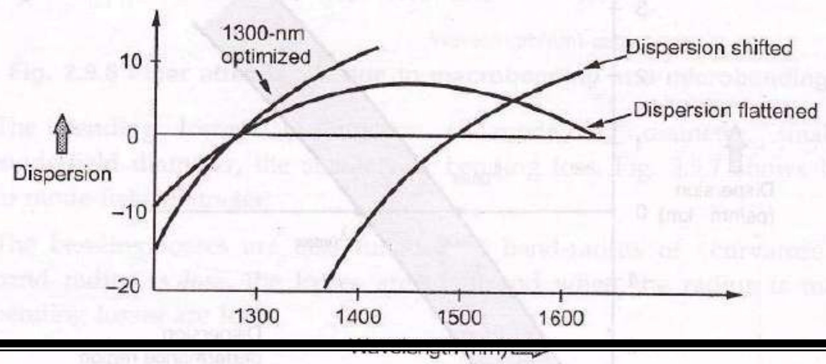


Fig. 2.9.4 Total resultant dispersion

Dispersion Calculations

The total dispersion consists of material and waveguide dispersions. The resultant intermodal dispersion is given as,

$$D(\lambda) = \frac{d\tau}{d\lambda}$$

where, τ is group delay per unit length of fiber.

The broadening σ of an optical pulse is given

$$\sigma = D(\lambda) L \sigma \lambda$$

where, $\sigma\lambda$ is half power spectral width of source.

As the dispersion varies with wavelength and fiber type. Different formulae are used to calculate dispersions for variety of fiber at different wavelength. For a non –

$$D(\lambda) = \frac{\lambda}{4} S_0 \left[1 - \left(\frac{\lambda_0}{\lambda}\right)^4\right]$$

dispersion shifted fiber between 1270 nm to 1340 nm wavelength, the expression for dispersion is given as :

where, λ_0 is zero dispersion wavelength.

S_0 is value at dispersion slop at λ_0.

Fig 2.9.5 shows dispersion performance curve for non-dispersion shifted fibers in 1270 – 1340 nm region.

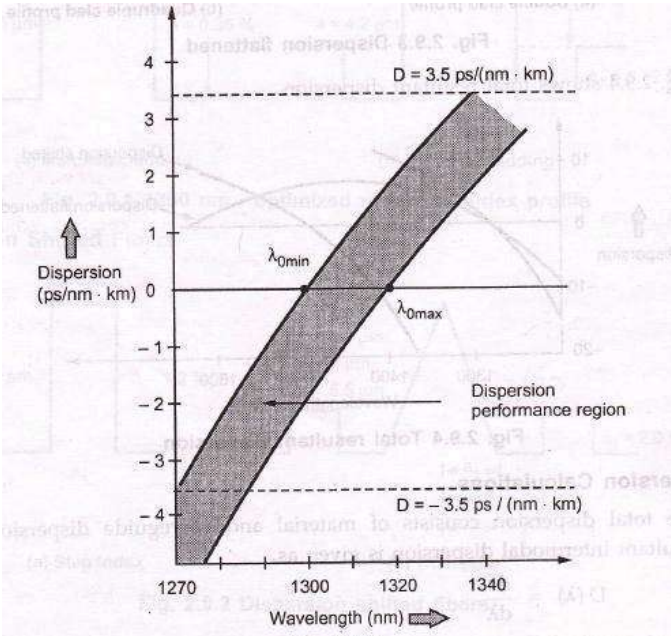


Fig. 2.9.5

Maximum dispersion specified as 3.5 ps/(nm . km) marked as dotted line in Fig. 2.9.5.

The cut-off frequency of an optical fiber

The cut-off frequency of an optical fiber is determined not only by the fiber itself (modal dispersion in case of multimode fibers and waveguide dispersion in case of single mode fibers) but also by the amount of material dispersion caused by the spectral width of transmitter.

Bending Loss Limitations

The macrobending and microbending losses are significant in single mode fibers at 1550 nm region, the lower cut-off wavelengths affects more. Fig. 2.9.6 shows macrobending losses.

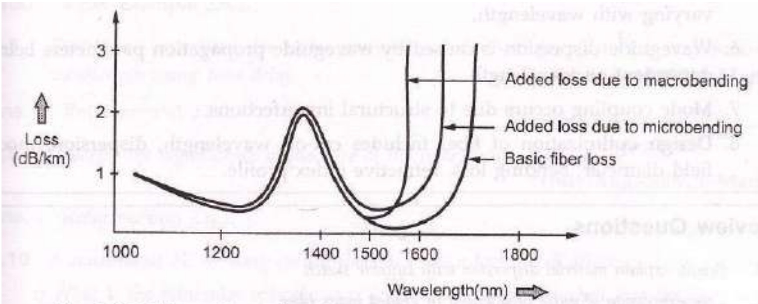


Fig. 2.9.6 Fiber attenuation due to macrobending and microbending

The bending losses are function of mode-filed diameter, smaller the mode-field diameter, the smaller the bending loss. Fig. 2.9.7 shows loss due to mode-field diameter.The bending losses are also function of bend-radius of curvature. If the bend radius is less, the losses are more and when the radius is more, the bending losses are less.

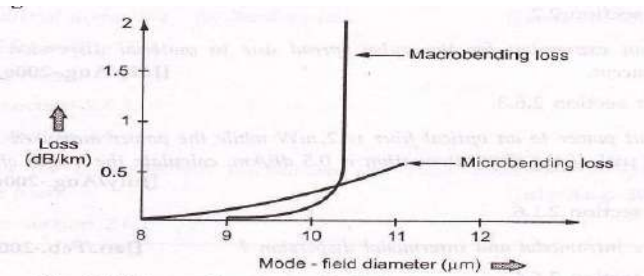


Fig. 2.9.7 Loss due to mode field diameter variation

UNIT III

OPTICAL SOURCES

3.1 Optical Sources

Optical transmitter coverts electrical input signal into corresponding optical signal. The

optical signal is then launched into the fiber. Optical source is the major component in an optical transmitter .Popularly used optical transmitters are Light Emitting Diode (LED) and semiconductor

Laser Diodes (LD)

Characteristics of Light Source of Communication

To be useful in an optical link, a light source needs the following characteristics

It must be possible to operate the device continuously at a variety of temperatures for many years.

- It must be possible to modulate the light output over a wide range of modulating frequencies. For fiber links, the wavelength of the output should coincide with one of transmission windows for the fiber type used.
- To couple large amount of power into an optical fiber, the emitting area should be small.
- To reduce material dispersion in an optical fiber link, the output spectrum should be narrow.
- The power requirement for its operation must be low.
- The light source must be compatible with the modern solid state devices.

- The optical output power must be directly modulated by varying the input current to the device.
- Better linearity of prevent harmonics and intermodulation distortion.
- High coupling efficiency.
- High optical output power.
- High reliability.
- Low weight and low cost.

Two types of light sources used in fiber optics are light emitting diodes (LEDs) and laser diodes (LDs).

Light Emitting Diodes(LEDs)

p-n Junction

Conventional p-n junction is called as **homojunction** as same semiconductor material is sued on both sides junction. The electron-hole recombination occurs in relatively layer = 10 μm. As the carriers are not confined to the immediate vicinity of junction, hence high current densities can not be realized.

The carrier confinement problem can be resolved by sandwiching a thin layer (= 0.1 μm) between p-type and n-type layers. The middle layer may or may not be doped. The carrier confinement occurs due to bandgap discontinuity of the junction. Such a junction is called heterojunction and the device is called double heterostructure.

In any optical communication system when the requirements is 1. Bit rate f 100-2—Mb/sec. 2. Optical power in tens of micro watts, LEDs are best suitable optical source.

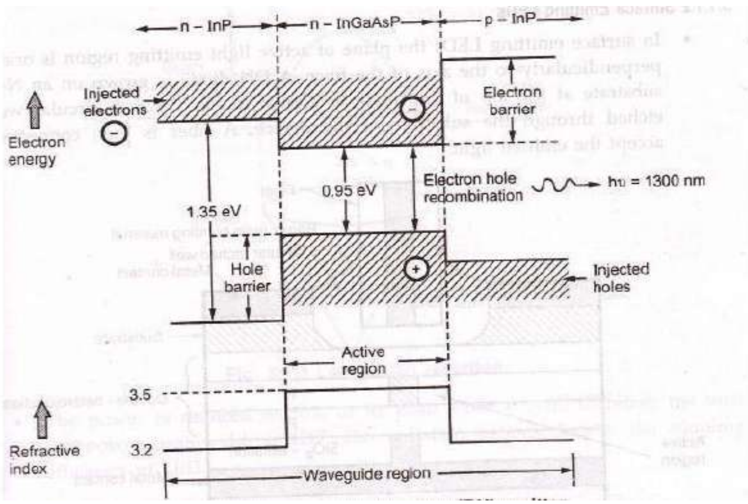
LED Structures

Heterojuncitons:

- A heterojunction is an interface between two adjoining single crystal semiconductors with different bandgap.
- Heterojuctions are of two types, Isotype (n-n or p-p) or Antisotype (p-n).

Double Heterojunctions (DH):

In order to achieve efficient confinement of emitted radiation double heterojunctions are used in LED structure. A heterojunciton is a junction formed by dissimilar semiconductors.Double heterojunction (DH) is formed by two different semiconductors on each side of active region. Fig. 3.1.1 shows double heterojunction (DH) light emitter.



The crosshatched regions represent the energy levels of free charge. Recombination occurs only in active In GaAsP layer. The two materials have different band gap energies and different refractive indices. The changes in band gap energies create potential barrier for both holes and electrons. The free charges can recombine only in narrow, well defined active layer side.

A double heterojunction (DH) structure will confine both hole and electrons to a narrow active layer. Under forward bias, there will be a large number of carriers injected into active region where they are efficiently confined. Carrier recombination occurs in small active region so leading to an efficient device. Another advantage DH structure is that the active region has a higher refractive index than the materials on either side, hence light emission occurs in an optical waveguide, which serves to narrow the output beam.

LED configurations

At present there are two main types of LED used in optical fiber links

Surface emitting LED

Edge emitting LED.

Both devices used a DH structure to constrain the carriers and the light to an active layer.

Surface Emitting LEDs

In surface emitting LEDs the plane of active light emitting region is oriented perpendicularly to the axis of the fiber. A DH diode is grown on an N-type substrate at the top of the diode as shown in Fig. 3.1.2. A circular well is etched through the substrate of the device. A fiber is then connected to accept the emitted

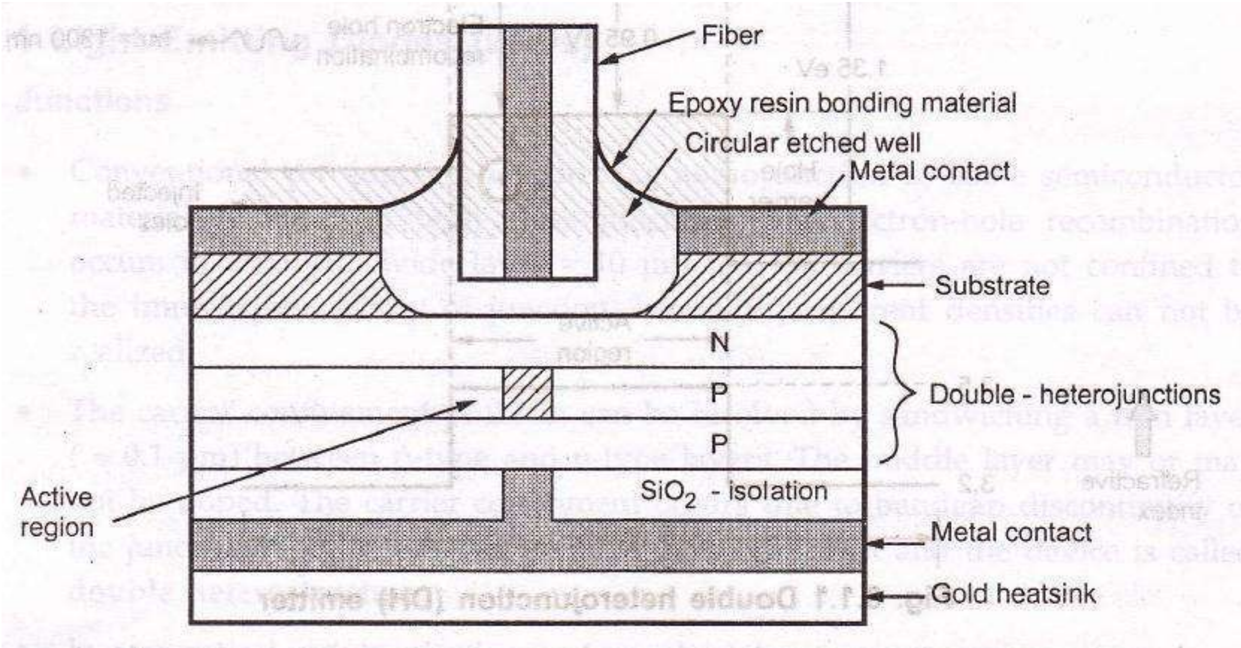


Fig. 3.1.2 Cross-section through a typical surface emitting LED

At the back of device is a gold heat sink. The current flows through the p-type material and forms the small circular active region resulting in the intense beam of light.

Diameter of circular active area = 50 µm

Thickness of circular active area = 2.5 µm

Current density = 2000 A/cm2 half-power

Emission pattern = Isotropic, 120o beamwidth.

The isotropic emission pattern from surface emitting LED is of Lambartian pattern. In Lambartian pattern, the emitting surface is uniformly bright, but its projected area diminishes as cos θ, where θ is the angle between the viewing direction and the normal to the surface as shown in Fig. 3.1.3. The beam intensity is maximum along the normal.

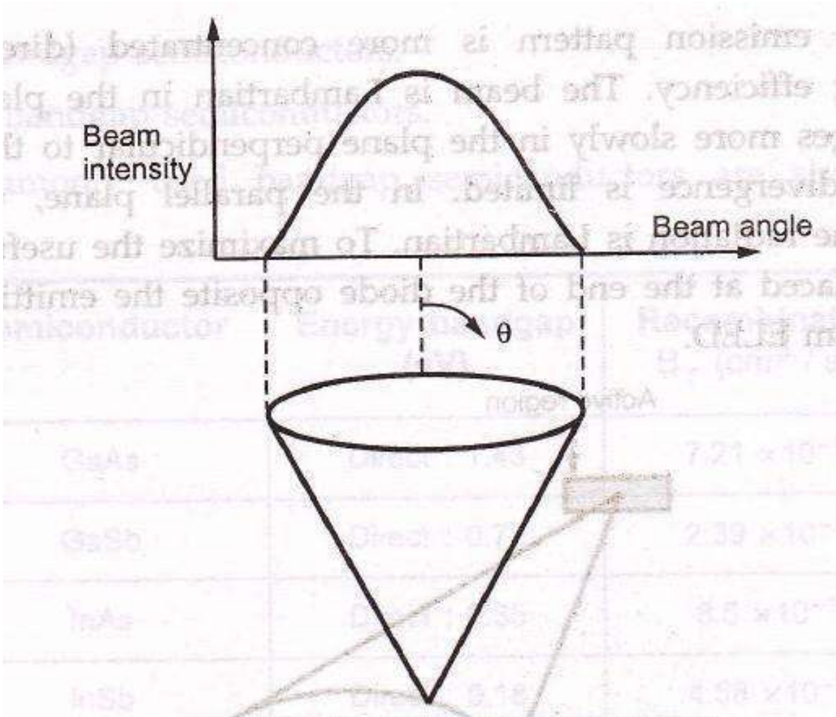


Fig. 3.1.3 Lambartian radiation

The power is reduced to 50% of its peak when θ = 60o, therefore the total half-power beamwidth is 120o. The radiation pattern decides the coupling efficiency of LED.

Edge Emitting LEDS (ELEDs)

In order to reduce the losses caused by absorption in the active layer and to make the beam more directional, the light is collected from the edge of the LED. Such a device is known as edge emitting LED or ELED.

It consists of an active junction region which is the source of incoherent light and two guiding layers. The refractive index of guiding layers is lower than active region but higher than outer surrounding material. Thus a waveguide channel is form and optical radiation is directed into the fiber. Fig. 3.1.4 shows structure of LED

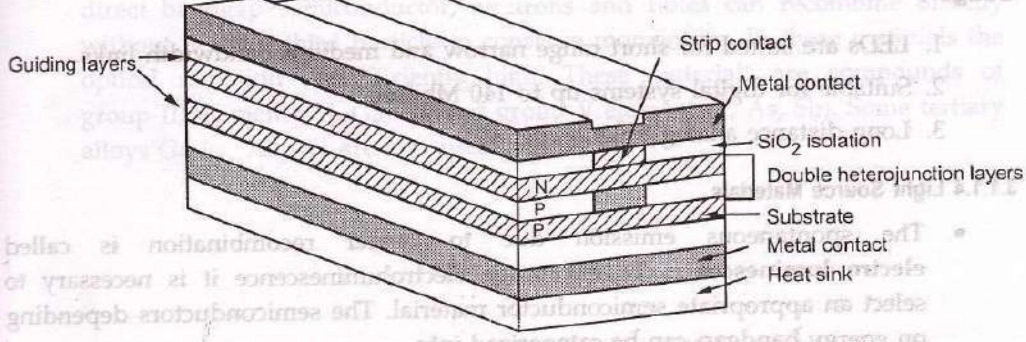


Fig. 3.1.4 Structure of edge emitting, DH, strip contact LED

Edge emitter's emission pattern is more concentrated (directional) providing improved coupling efficiency. The beam is Lambartian in the plane parallel to the junction but diverges more slowly in the plane perpendicular to the junction. In this plane, the beam divergence is limited. In the parallel plane, there is no beam confinement and the radiation is Lambartian. To maximize the useful output power, a reflector may be placed at the end of the diode opposite the emitting edge. Fig. 3.1.5 shows radiation from ELED.

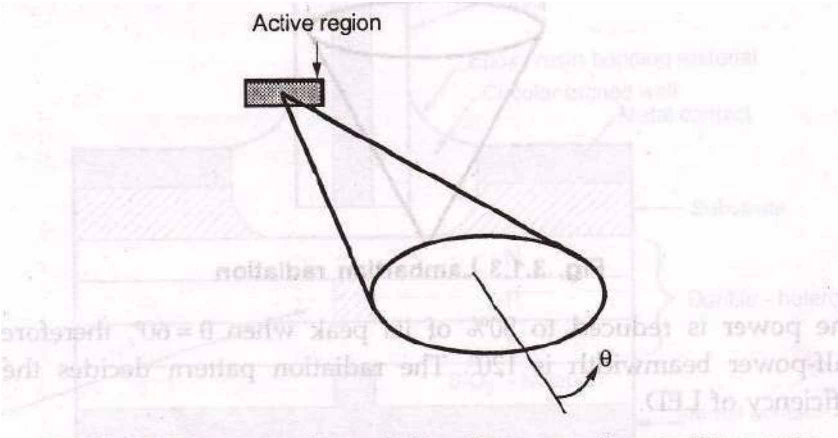


Fig. 3.1.5 Unsymmetric radiation from an edge emitting LED

Features of ELED:

- Linear relationship between optical output and current.
- Spectral width is 25 to 400 nm for $\lambda = 0.8 - 0.9\ \mu m$.
- Modulation bandwidth is much large.
- Not affected by catastrophic gradation mechanisms hence are more reliable.
- ELEDs have better coupling efficiency than surface emitter.
- ELEDs are temperature sensitive.

Usage :

1. LEDs are suited for short range narrow and medium bandwidth links.
2. Suitable for digital systems up to 140 Mb/sec.
3. Long distance analog links

Light Source Materials

The spontaneous emission due to carrier recombination is called electro luminescence. To encourage electroluminescence it is necessary to select as appropriate semiconductor material. The semiconductors depending on energy bandgap can be categorized into

Direct bandgap semiconductors.

Indirect bandgap semiconductors.

Some commonly used bandgap semiconductors are shown in following table 3.1.1

Semiconductor	**Energy bandgap (eV)**	**Recombination Br (cm3 / sec)**
GaAs	Direct : 1.43	7.21 x 10-10
GaAs	Direct : 0.73	2.39 x 10-10
InAs	Direct : 0.35	8.5 x 10-11
InSb	Direct : 0.18	4.58 x 10-11
Si	Indirect : 1.12	1.79 x 10-15
Ge	Indirect : 0.67	5.25 x 10-14
GaP	Indirect : 2.26	5.37 x 10-14

Table 3.1.1 Semiconductor material for optical sources

Direct bandgap semiconductors are most useful for this purpose. In direct bandgap semiconductors the electrons and holes on either side of bandgap have same value of crystal momentum. Hence direct recombination is possible. The recombination occurs within 10-8 to 10-10 sec.

In indirect bandgap semiconductors, the maximum and minimum energies occur at different values of crystal momentum. The recombination in these semiconductors is quite slow i.e. 10^{-2} and 10^{-3} sec.

The active layer semiconductor material must have a direct bandgap. In direct bandgap semiconductor, electrons and holes can recombine directly without need of third particle to conserve momentum. In these materials the optical radiation is sufficiently high. These materials are compounds of group III elements (Al, Ga, In) and group V element (P, As, Sb). Some tertiary allos Ga1-x Alx As are also used.

Emission spectrum of Ga1-x AlxAs LED is shown in Fig. 3.1.6.

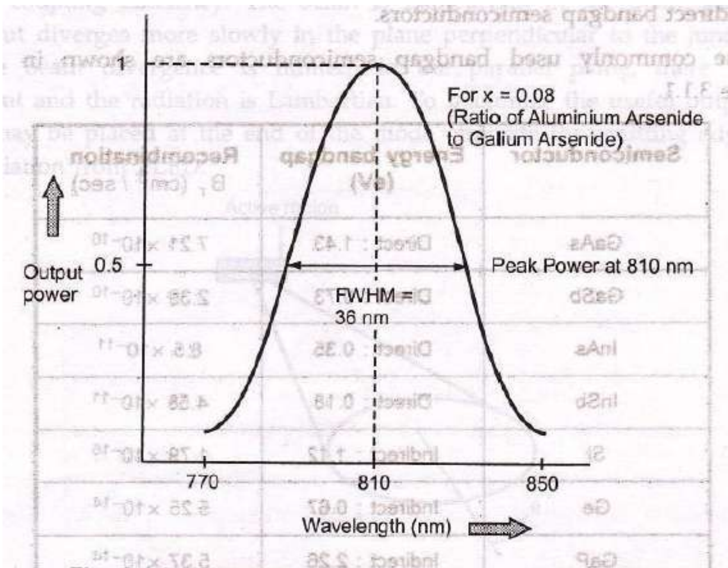


Fig. 3.1.6 Emission spectrum of $Ga_{1-x}\ Al_x As$ LED

The peak output power is obtained at 810 nm. The width of emission spectrum at half power (0.5) is referred as full width half maximum (FWHM) spectral width. For the given LED FWHM is 36 nm.

The fundamental quantum mechanical relationship between gap energy E and frequency

v is given as

$$E = hv$$

$$E = h\frac{c}{\lambda}$$

$$\lambda = \frac{hc}{E}$$

where, energy (E) is in joules and wavelength (λ) is in meters. Expressing the gap energy (Eg) in electron volts and wavelength (λ) in micrometers for this application.

$$\lambda(\mu m) = \frac{1.24}{E_g(eV)}$$

Different materials and alloys have different band gap energies

The bandgap energy (Eg) can be controlled by two compositional parameters x and y, within direct bandgap region. The quartenary alloy In1-x Gax Asy P1-y is the principal material sued in such LEDs. Two expression relating Eg and x,y are –

$$E_g = 1.424 + 1.266\,x + 0266\,x^2$$

$$E_g = 1.35 - 0.72\,y + 0.12\,y^2$$

Example 3.1.1 : Compute the emitted wavelength from an optical source having x = 0.07.

Solution : x = 0.07

$$E_g = 1.424 + 1.266\,x + 0266\,x^2$$

$$E_g = 1.424 + (1.266 \times 0.07) + 0266 \times (0.07)^2$$

Eg = 1.513 eV

now

$$\boxed{\lambda = \frac{1.24}{E_g}}$$

$$\lambda = \frac{1.24}{1.513}$$

$$\lambda = 0.819\ \mu m$$

$$\lambda = \mathbf{0.82}\ \mu m$$

Example 3.1.2 : For an alloy In0.74 Ga0.26 As0.57 P0.43 to be sued in Led. Find the wavelength emitted by this source.

Solution : Comparing the alloy with the quartenary alloy composition. In1-x Gax As P1-y it is found that

x = 0.26 and y= 0.57

Eg = 1.35 – 0.72 y + 0.12 y2

Using

Eg = 1.35-(0.72 x 0.57) + 0.12 x 0.572

Eg = 0.978 eV

now

$$\lambda = \frac{1.24}{E_g}$$

$$\lambda = \frac{1.24}{0.978}$$

$$\lambda = 1.2671\ \mu m$$

$$\lambda = \mathbf{1.27}\ \mu\mathbf{m}$$

Quantum Efficiency and Power

The internal quantum efficiency (η_{int}) is defined as the ratio of radiative recombination rate to the total recombination rate.

$$\eta_{int} = \frac{R_r}{R_r + R_{nr}}$$

Where,

R_r is radiative recombination rate.

R_{nr} is non-radiative recombination rate.

$$\tau_r = \frac{n}{R_r}$$

If n are the excess carriers, then radiative life time, and non-radiative life time,

$$\tau_r = \frac{n}{R_{nr}}$$

The internal quantum efficiency is given The recombination time of carriers in active region is τ. It is also known as bulk recombination life time.

$$\eta_{int} = \frac{1}{1+\frac{R_{nr}}{R_r}}$$

$$\boxed{\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}}}$$

$$\eta_{int} = \frac{1}{1+\frac{\tau_r}{\tau_{nr}}}$$

Therefore internal quantum efficiency is given as –

$$\boxed{\eta_{int} = \frac{\tau}{\tau_r}}$$

If the current injected into the LED is I and q is electron charge then total number of re combinations per second is –

$$R_r = R_{nr} = \frac{I}{q}$$

$$\eta_{int} = \frac{R_r}{I / q}$$

$$R_r = \eta_{int} \times \frac{I}{q}$$

Optical power generated internally in LED is given as –

$$P_{int} = R_r . h\,v$$

$$P_{int} = \left(\eta_{int} \times \frac{I}{q}\right) . h\,v$$

$$P_{int} = \left(\eta_{int} \times \frac{I}{q}\right) . h \frac{c}{\lambda}$$

$$P_{int} = \eta_{int} \cdot \frac{hc\,I}{q\lambda}$$

Not all internally generated photons will available from output of device. The external quantum efficiency is used to calculate the emitted power. The external quantum efficiency is defined as the ratio of photons emitted from LED to the number of photons generated internally. It is given by equation

$$\eta_{ext} = \frac{1}{n(n+1)^2}$$

The optical output power emitted from LED is given as –

$$P = \eta_{ext} \cdot P_{int}$$

$$P = \frac{1}{n\,(n+1)^2} \cdot P_{int}$$

Example 3.1.3 : The radiative and non radiative recombination life times of minority carriers in the active region of a double heterojunction LED are 60 nsec and 90 nsec respectively. Determine the total carrier recombination life time and optical power generated internally if the peak emission wavelength si 870 nm and the drive currect is 40 mA.

Solutions:

Given : λ = 870 nm 0.87 x 10-6 m

τr = 60 nsec.

τnr = 90 nsec.

I = 40 mA = 0.04 Amp.

i) Total carrier recombination life time:

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}}$$

$$\frac{1}{\tau} = \frac{1}{60} + \frac{1}{90}$$

$$\frac{1}{\tau} = \frac{150}{5400}$$

ii) Internal optical power

iii)

$$P_{int} = \eta_{int} \cdot \frac{hc\,I}{q\lambda}$$

iv)

$$P_{int} = \left(\frac{\tau}{\tau_r}\right)\left(\frac{hc\,I}{q\lambda}\right)$$

$$P_{int} = \left(\frac{30}{60}\right)\left[\frac{(6.625 \times 10^{-34})(3 \times 10^{8}) \times 0.04}{(1.602 \times 10^{-19})(0.87 \times 10^{-6})}\right]$$

Example 3.1.4 : A double heterjunciton InGaAsP LED operating at 1310 nm has radiative and non-radiative recombination times of 30 and 100 ns respectively. The current injected is 40 Ma.

Calculate – Bulk recombination life time. Internal quantum efficiency Internal power level.

Solution : λ = 1310 nm = (1.31 x 10-6 m)

τr = 30 ns

τnr = 100 ns

i) I = 40 MA – 0.04 Amp.

Bulk Recombination Life time (τ) :

$$\eta_{int} = \frac{\tau}{\tau_r}$$

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}}$$

Internal quantum efficienty (ηint)

$$\eta_{int} = \frac{23.07}{30}$$

$$\eta_{int} = \mathbf{0.769}$$

Internal power level (Pint) :

$$P_{int} = \eta_{int} . \frac{hc\, I}{q\lambda}$$

Advantages and Disadvantages of LED

Advantages of LED

ii) Simple design.

iii) Ease of manufacture.

iv) Simple system integration.

v) Low cost.

vi) High reliability.

Disadvantages of LED

1. Refraction of light at semiconductor/air interface.
2. The average life time of a radiative recombination is only a few nanoseconds, therefore
3. Modulation BW is limited to only few hundred megahertz.
4. Low coupling efficiency.
5. Large chromatic dispersion.

Comparison of Surface and Edge Emitting LED

LED type	Max. modulation freq. (MHz)	Output power (mW)	Fiber coupled power
Surface emitting	60	< 4	< 0.2
Edge emitting	200	< 7	< 1.0

Injection Laser Diode (ILD)

The laser is a device which amplifies the light, hence the LASER is an acronym for light amplification by stimulated emission of radiation.

The operation of the device may be described by the formation of an electromagnetic standing wave within a cavity (optical resonator) which provides an output of monochromatic highly coherent radiation.

Principle :

Material absorb light than emitting. Three different fundamental process occurs between the two energy □□ states of an atom.□ Absorption 2) Spontaneous emission 3) Stimulated emission.Laser action is the result of three process absorption of energy packets (photons) spontaneous emission, and stimulated emission. (These processes are represented by the simple two-energy-level diagrams). Where E1 is the lower state energy level. E2 is the higher state energy level.

Quantum theory states that any atom exists only in certain discrete energy state, absorption or emission of light causes them to make a transition from one state to another. The frequency of the absorbed or emitted radiation f is related to the difference in energy E between the two states. If E1 is lower state energy level. and E2 is higher state energy level

E = (E2 – E1) = h.f. Where, h = 6.626 x 10-34 □□/s (Plank's constant).

An atom is initially in the lower energy state, when the photon with energy (E2 – E1) is incident on the atom it will be excited into the higher energy state E2 through the absorption of the photon

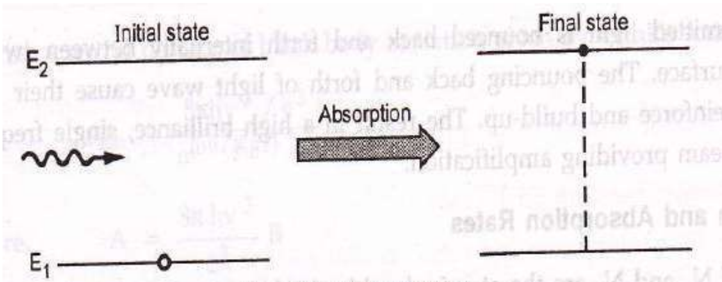


Fig. 3.1.7 Absorption

When the atom is initially in the higher energy state E2, it can make a transition to the lower energy state E1 providing the emission of a photon at a frequency corresponding to $\Delta E = h.f.$ The emission process can occur in two ways. By spontaneous emission in which the atom returns to the lower energy state in random manner.

By stimulated emission when a photon having equal energy to the difference between the two states (E2 – E1) interacts with the atom causing it to the lower state with the creation of the second photon

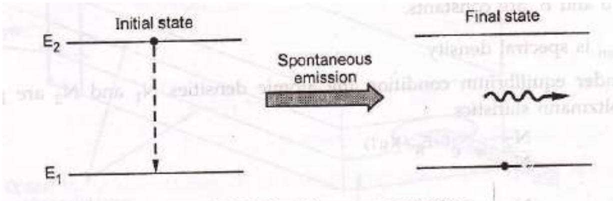


Fig. 3.1.8 Spontaneous emission

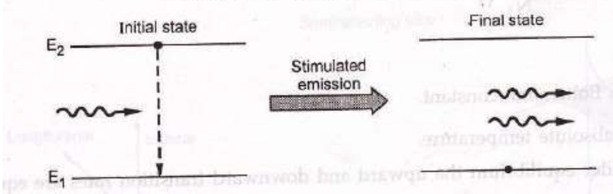


Fig. 3.1.9 Stimulated emission

Spontaneous emission gives incoherent radiation while stimulated emission gives coherent radiation. Hence the light associated with emitted photon is of same frequency of incident photon, and in same phase with same polarization.

It means that when an atom is stimulated to emit light energy by an incident wave, the liberated energy can add to the wave in constructive manner. The emitted light is bounced back and forth internally between two reflecting surface. The bouncing back and forth of light wave cause their intensity to reinforce and build-up. The result in a high brilliance, single frequency light beam providing amplification.

Emission and Absorption Rates

It N1 and N2 are the atomic densities in the ground and excited states.

Rate of spontaneous emission

Rspon = AN2

Rate of stimulated emission

Rstim = BN2 ρem

Rate of absorption

Rabs = B' N1 ρem

where,

A, B and B' are constants.

ρem is spectral density.

Under equilibrium condition the atomic densities N1 and N2 are given by Boltzmann statistics.

$$\frac{N_2}{N_1} = eg^{(-E_B / K_B T)}$$

$$\frac{N_2}{N_1} = eg^{(-h_v / K_B T)}$$

where,

KB is Boltzmann constant.

T is absolute temperature.

Under equilibrium the upward and downward transition rates are equal.

AN2 + BN2 ρem = B' N1 ρem

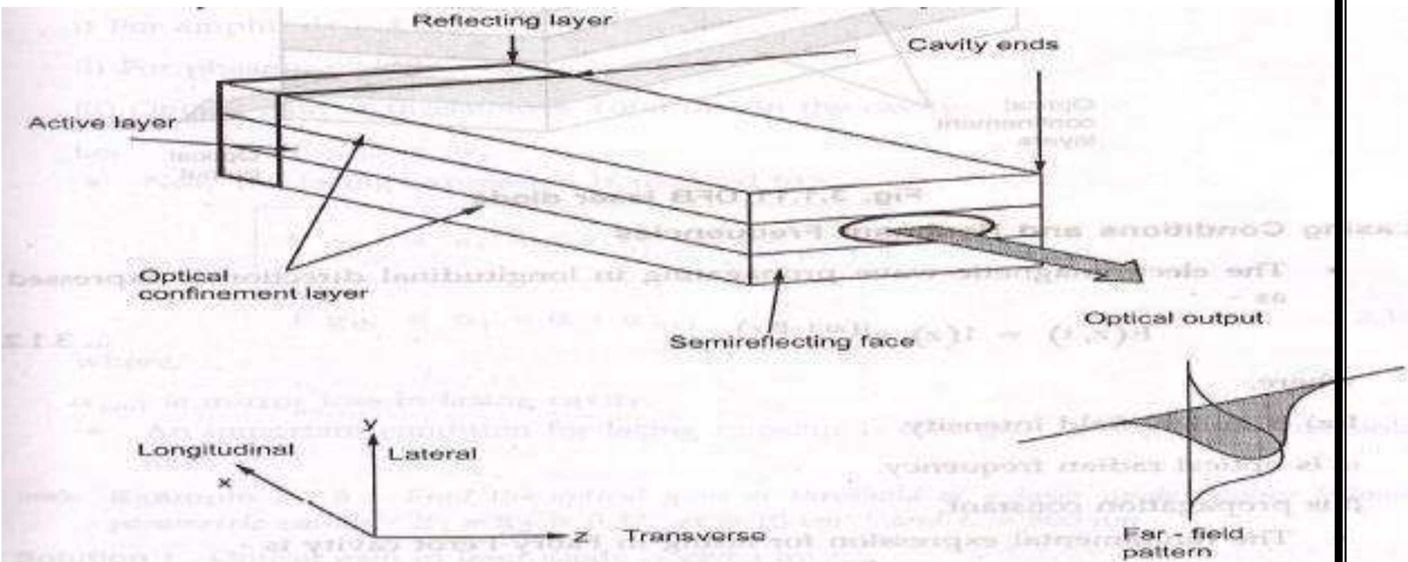
Spectral density pem

Comparing spectral density of black body radiation given by Plank's formula,

Therefore, … A and B are called Einstein's coefficient.

Fabry – Perot Resonator

Lasers are oscillators operating at frequency. The oscillator is formed by a resonant cavity providing a selective feedback. The cavity is normally a Fabry-Perot resonator i.e. two parallel plane mirrors separated by distance L,



Light propagating along the axis of the interferometer is reflected by the mirrors back to the amplifying medium providing optical gain. The dimensions of cavity are 25-500 µm longitudinal 5-15 µm lateral and 0.1-0.2 µm transverse. Fig. 3.1.10 shows Fabry-Perot resonator cavity for a laser diode.

The two heterojunctions provide carrier and optical confinement in a direction normal to the junction. The current at which lasing starts is the threshold current. Above this current the output power increases sharply.

Distributed Feedback (DFB) Laser

In DFB laster the lasing action is obtained by periodic variations of refractive index along the longitudinal dimension of the diode. Fig. 3.1.11 shows the structure of DFB laser diode

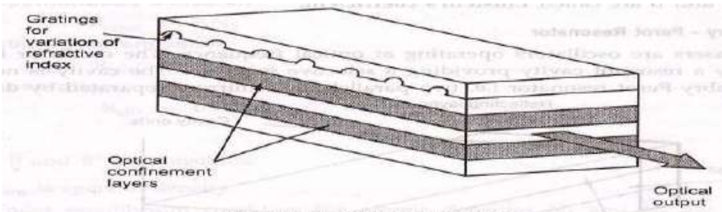


Fig. 3.1.11 DFB laser diode

Lasing conditions and resonant Frequencies

The electromagnetic wave propagating in longitudinal direction is expressed as –

E(z, t) = I(z) ej(ω t-β z)

where,

I(z) is optical field intensity.

is optical radian frequency.

β is propagation constant.

The fundamental expression for lasing in Fabry-Perot cavity is –

$$I(z) = I(0)e^{[(\Gamma g(h\nu) - \alpha(h\nu))z]}$$

where, is optical field confinement factor or the fraction of optical power in the active layer.

α is effective absorption coefficient of material.

g is gain coefficient.

h v is photon energy.

z is distance traverses along the lasing cavity.

The condition of lasing threshold is given as –

For amplitude : I (2L) = I (0)

For phase : e-j2β L = 1

Optical gain at threshold = Total loss in the cavity.

i.e. Γ gth = αt

Now the lasing expression is reduced to –

$$\Gamma g_{th} = a_t = \propto + \frac{1}{2L} \ln\left(\frac{1}{R_1 R_2}\right)$$

$$\Gamma g_{th} = \propto_t = \propto + \propto_{end}$$

where,

Aend is mirror loss in lasing cavity. An important condition for lasing to occur is that gain, g ≥ g th i.e. threshold gain.

Example 3.1.5 : Find the optical gain at threshold of a laser diode having following parametricvalues – R1 = R2 = 0.32, α = 10cm-1 and L = 500 µm.

Solution : Optical gain in laser diode is given by –

$$\Gamma g_{th} = 10 + \frac{1}{2 \times (500 \times 10^{-4})} \ln\left(\frac{1}{0.32 \times 0.32}\right)$$

$$\Gamma g_{th} = 33.7\ \text{cm}^{-1}$$

Power Current Characteristics

The output optic power versus forward input current characteristics is plotted in Fig. 3.1.12 for a typical laser diode. Below the threshold current (Ith) only spontaneous emission is emitted hence there is small increase in optic power with drive current. At threshold when lasing conditions are satisfied. The optical power increases sharply after the lasing threshold because of stimulated emission.The lasing threshold optical gain (gth) is related by threshold current density (Jth) for stimulated emission by expression –

g th = β

where, β is constant for device structure

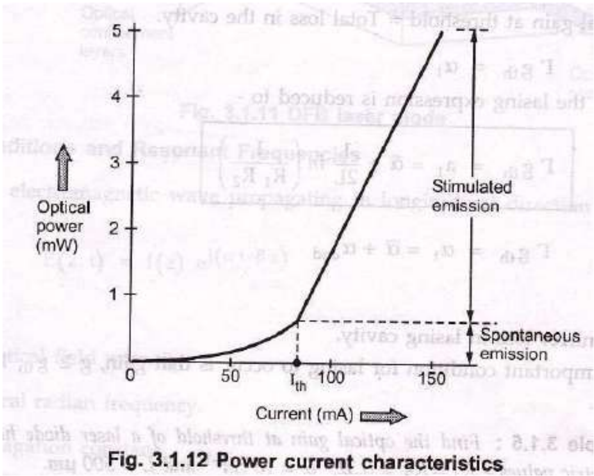


Fig. 3.1.12 Power current characteristics

External Quantum Efficiency

The external quantum efficiency is defined as the number of photons emitted per electron hole pair recombination above threshold point. The external quantum efficiency ηext is given by –

$$\eta_{ext} = \frac{\eta_i (g_{th} - \alpha)}{g_{th}}$$

where,

ηi = Internal quantum efficiency (0.6-0.7).

gth = Threshold gain.

α = Absorption coefficient

Typical value of ηext for standard semiconductor laser is ranging between 15-20 %.

Resonant Frequencies

At threshold lasing

2β L = 2π m

where, (propagation constant)

m is an integer.

$$= 2L.\frac{n}{\lambda}$$

Since c = vλ

$$\lambda = \frac{c}{v}$$

Substituting λ in 3.1.30

$$m = 2L\,\frac{nv}{c}$$

Gain in any laser is a function of frequency. For a Gaussian output the gain and frequency are related by expression –

$$g(\lambda) = g(0)e^{\left[-\frac{(\lambda-\lambda_0)^2}{2\sigma^2}\right]}$$

where, g(0) is maximum gain. λ_0 is center wavelength in spectrum.is spectral width of the gain.The frequency spacing between the two successive modes is –

$$\Delta v = \frac{c}{2Ln}$$

$$\Delta\lambda = \frac{\lambda^2}{2Ln}$$

Optical Characteristics of LED and Laser

The output of laser diode depends on the drive current passing through it. At low drive current, the laser operates as an inefficient Led, When drive current crosses threshold value, lasing action beings. Fig. 3.1.13 illustrates graph comparing optical powers of LED operation (due to spontaneous emission) and laser operation (due to stimulated emission).

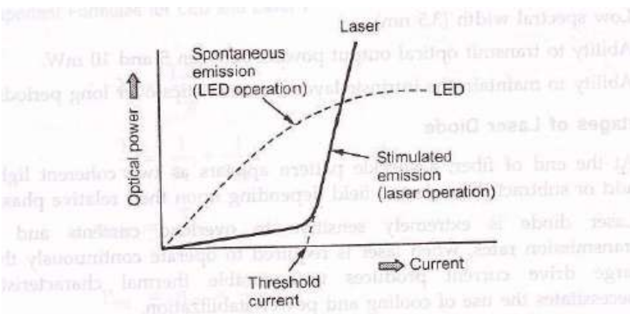


Fig. 3.1.13 Optical characteristics of an LED and laser compared

Spectral and Spatial Distribution of Led and Laser

At low current laser diode acts like normal LED above threshold current, stimulated emission i.e. narrowing of light ray to a few spectral lines instead of broad spectral distribution, exist. This enables the laser to easily couple to single mode fiber and reduces the amount of uncoupled light (i.e. spatial radiation distribution). Fig. 3.1.14 shows spectral and spatial distribution difference between two diodes

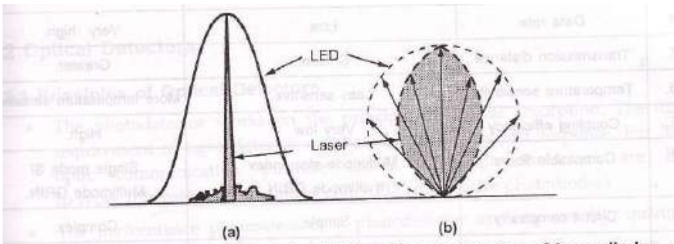


Fig. 3.1.14 Comparing (a) spectral and (b) spatial distribution of laser diodes

Advantages and Disadvantages of Laser Diode

Advantages of Laser Diode

- Simple economic design.
- High optical power.
- Production of light can be precisely controlled.
- Can be used at high temperatures.

- Better modulation capability.
- High coupling efficiency.
- Low spectral width (3.5 nm)
- Ability to transmit optical output powers between 5 and 10 mW.
- Ability to maintain the intrinsic layer characteristics over long periods.

Disadvantages of Laser Diode

- At the end of fiber, a speckle pattern appears as two coherent light beams add or subtract their electric field depending upon their relative phases.
- Laser diode is extremely sensitive to overload currents and at high transmission rates, when laser is required to operate continuously the use of large drive current produces unfavourable thermal characteristics and necessitates the use of cooling and power stabilization.

Comparison of LED and Laser Diode

S. No.	Parameter	LED	LD (Laser Diode)
1	Principle of operation	Spontaneous emission.	Stimulated emission.
2	Output beam	Non – coherent	Coherent.
3	Spectral width	Board spectrum (20 nm – 100 nm)	Much narrower (1-5 nm).
4	Data rate	Low.	Very high.
5	Transmission distance	Smaller.	Greater.
6	Temperature sensitivity	Less sensitive.	More temperature

			sensitive
7	Coupling efficiency	Very low.	High.
8	Compatible fibers	Multimode step index multimode GRIN	Single mode SI Multimode GRIN
9	Circuit complexity	Simple	Complex
10	Life time	105 hours.	104 hours.
11	Cost	Low.	High.
12	Output power	Linearly proportional to drive current	Proportional to current above the threshold
13	Current required	Drive current 50 to 100 mA peak.	Threshold current 5 to 40 ma
14	Applications	Moderate distance low data rate.	Long distance high data rates.

Important Formulae for LED and Laser

LED

- $\lambda = \frac{1.24}{E_g}$
- $\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}}$
- $\eta_{int} = \frac{\tau}{\tau_r}$

- $P_{int} = \eta_{int} \times \frac{hc\,I}{q\lambda}$

LASER

1. $\Gamma g_{th} = \alpha + \frac{1}{2L} \ln\left(\frac{1}{R_1 R_2}\right)$
2. $\Delta v = \frac{c}{2\,L\,n}$
3. $\Delta \lambda = \frac{\lambda^2}{2\,L\,n}$

UNIT IV

OPTICAL DETECTORS AND RECEIVERS

- The photo-diode is in fact a p-n junction put to the exact opposite use as the LED
- The variation in current is a function of the incident light
- Use of the stimulated absorption of light by the semiconductor material for the generation of electron-hole pairs.
- The energy of the absorbed photons to transfer the electrons from the ground to the excited state contributes to the variation in circuit current.
- The energy of the absorbed photon must at least be equal to the band-gap of the material for the material to respond to the incoming photons.

PIN diode

- A simple way to increase the depletion-region width is to insert a layer of undoped (or lightly doped) semiconductor material between the p–n junction.
- Since the middle layer consists of nearly intrinsic material, such a structure is referred to as the p–i–n photodiode.

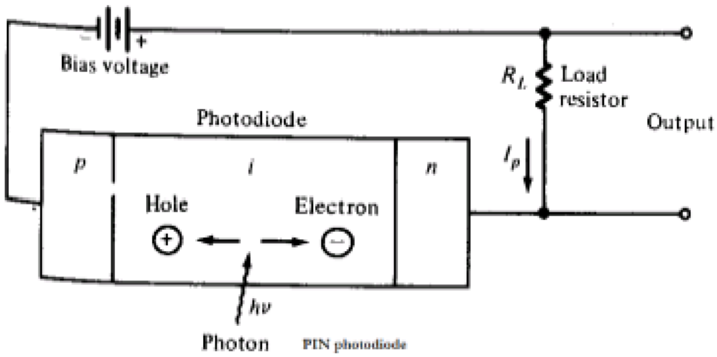


Figure 3.11

- When photon enters photodetector, the low band gap absorption layer absorbs the photon, and an electron-hole pair is generated. This electron hole pair is called photocarrier.
- These photocarriers, under the influence of a strong electric field generated by a reverse bias potential difference across the device as shown in figure produce photocurrent proportional to number of incident photons.

Avalanche Photo Diode (APD)

- All detectors require a certain minimum current to operate reliably. The current requirement translates into a minimum power requirement through Pin=IpRPin=IpR.
- Detectors with a large responsivity R are preferred since they require less optical power.
- The responsivity of p–i–n photodiodes is limited while Avalanche photodiode (APDs) can have much larger values of R.

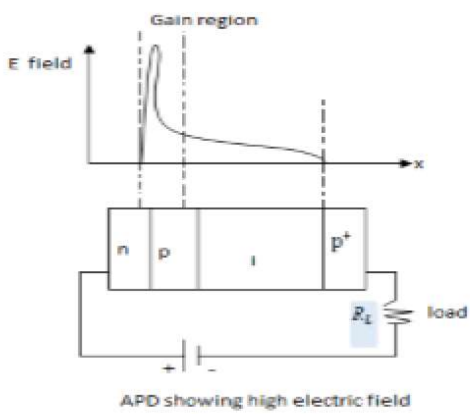


Figure 3.12: APD

Working of APD

- APD is similar to PIN diode the exception is the addition of high intensity electric field region.
- In this region primary electron hole pairs are generated by the incident photons which are able to absorb enough kinetic energy from strong electric field to collide with the atoms present in this region, thus generating more electron hole pairs.
- The physical phenomenon behind the internal current gain is known as the impact ionization.
- This impact ionization leads to avalanche breakdown in ordinary reverse bias. It requires very high reverse bias voltage in order that the new carriers created by impact ionization can themselves produce additional carriers by same mechanism.
- This process of generating more than one electron hole pair from incident photon through ionization process is referrednto as the avalanche effect.
- Thus the avalanche multiplication results in amplification of photodiode current.
- Multiplication factor: Multiplication factor M is a measure of internal gain provided by APD. It is defined as the ratio of total multiplied output current to the primary un multiplied current.

M=IIpM=IIp

Where I the Total multiplied output current

IpIp is the primary un multiplied current

Multiplication depends on physical and operational characteristics of photo detector device. Operational characteristics include the width of avalanche region, the strength of electric field and type of semiconductor material employed.

Comparison:

Sr no.	**PIN diode**	**APD (Avalanche photodiode)**
1	PIN does not have high intensity electric field region.	APD has high intensity electric field region.
2	Photo current (IpIp) generated is less compared to APD Ip=qNθ, q = electron charge, Nθ = carrier number	Photo current (IpIp) generated is more compared to PIN, Ip=qNθ.M q = electron charge, Nθ= carrier number, M = multiplication factor
3	Responsively of PIN is limited.	Responsively of APD can have much larger values.
4	They exhibit lower noise levels.	They exhibit higher noise levels as compared to PIN due to impact ionization and photocurrent multiplication.
5	Response time of PIN is half that of APD.	Response time of APD is almost double that of PIN.

The PIN-diode is an alteration of the PN-junction for particular applications. After the PN-junction diode was developed in the year 1940s, the diode was first exercised as a high-power rectifier, low-frequency during the year 1952. The occurrence of an intrinsic layer can significantly increase the breakdown voltage for the application of high-voltage. This intrinsic layer also offers exciting properties when the device operates at high frequencies in the range of radio wave and microwave. A PIN diode is a one kind of diode with an undoped, wide intrinsic semiconductor region between a P-type and N-type semiconductor region. These regions are normally heavily doped as they are used for Ohmic contacts. The wider intrinsic region is indifference to an ordinary p–n diode. This region makes the diode an inferior rectifier but it makes it appropriate for fast switches, attenuators, photo detectors and high voltage power electronics applications.

The PIN diode is a one type of photo detector, used to convert optical signal into an electrical signal. The PIN diode comprises of three regions, namely P-region, I-region and N-region. Typically, both the P and N regions are heavily doped due to they are utilized for Ohmic contacts. The intrinsic region in the diode is in contrast to a PN junction diode. This region makes the PIN diode a lower rectifier, but it makes it appropriate for fast switches, attenuators, photo detectors and applications of high voltage power electronics.

Photo detectors:

These are Opto-electric devices i.e. to convert the optical signal back into electrical impulses. The light detectors are commonly made up of semiconductor material. Photo detectors made up of semiconductor material. When the light strikes the light detector a current is produced in the external circuit proportional to the intensity of the incident light.

Optical signal generally is weakened and distorted when it emerges from the end of the fiber, the photo detector must meet following strict performance requirements.

- A high sensitivity to the emission wavelength range of the received light signal
- A minimum addition of noise to the signal
- A fast response speed to handle the desired data rate
- Be insensitive to temperature variations
- Be compatible with the physical dimensions of the fiber
- Have a Reasonable cost compared to other system components
- Have a long operating lifetime

Some important parameters while discussing photo detectors:

Quantum Efficiency

It is the ratio of primary electron-hole pairs created by incident photon to the photon incident on the diode material.

Detector Responsivity

This is the ratio of output current to input optical power.Hence this is the efficiency of the device.

Spectral Response Range

This is the range of wavelengths over which the device will operate.

Noise Characteristics

The level of noise produced in the device is critical to its operation at low levels of input light.

Response Time

This is a measure of how quickly the detector can respond to variations in the input light intensity.

Types of Light Detectors

- PIN Photodiode

- Avalanche Photodiode

PIN photodiode

InGaAs avalanche photodiode

Photo detector materials

Operating Wavelength Ranges for Several Different Photo detector Materials

Material	Energy gap, eV	λ_{cutoff}, nm	Wavelength range, nm
Silicon	1.17	1060	400–1060
Germanium	0.775	1600	600–1600
GaAs	1.424	870	650–870
InGaAs	0.73	1700	900–1700
InGaAsP	0.75–1.35	1650–920	800–1650

In GaAs is used most commonly for both long-wavelength pin and avalanche photodiodes

Physical Principles of Photodiodes

The Pin Photodetector

- The device structure consists of p and n semiconductorregions separated by a very lightly n-doped intrinsic (i) region.
- In normal operation a reverse-bias voltage is applied across the device so that no free electrons or holes exist in the intrinsic region.
- Incident photon having energy greater than or equal to the band gap energy of the semiconductor material, give up its energy and excite an electron from the valence band to the conduction band

Pin Photodetector

the high electric field present in the depletion region causes photogenerated carriers to separate and be collected across the reverse – biased junction. This gives rise to a current flow in an external circuit, known as photocurrent.

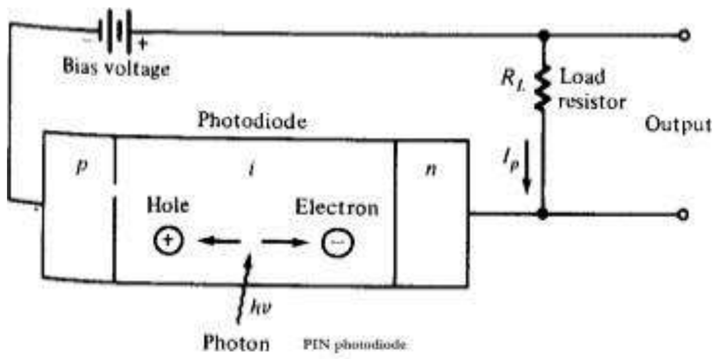


Figure 3.11

Module -3 Notes

Mobile Communication Engineering: Wireless Network generations, Basic propagation Mechanisms, Mobile radio Channel. Principles of Cellular Communications: Cellular terminology, Cell structure and Cluster, Frequency reuse concept, Cluster size and system capacity, Frequency Reuse Distance, Co-channel Interference and signal quality

1.4 WIRELESS NETWORK GENERATIONS

The cellular systems have been classified into three distinct evolutions of generations:

Q.1 Explain the 1G Analog Cellular system

1.4.1 First-Generation Analog Cellular Systems

- The first-generation (1G) analog cellular communication systems are voice-oriented analog cellular systems using frequency division multiple access technique.
- The first-generation systems used large cells and omni-directional antennas in the 800-MHz band.
- The AMPS and ETACS systems use a seven-cell reuse pattern with provisions for cell-sectoring and cell-splitting to increase capacity when needed.
- Limited traffic-handling capacity.
- The first-generation cellular systems are based on analog transmission technology.
- The most popular first generation cellular systems are AMPS (widely deployed in most parts of US, South America, Australia, China), and ETACS (deployed throughout Europe).
- The systems transmit speech signals employing FM, and
- Important control information is transmitted in digital form using FSK.
- The entire service area is divided into logical cells, and each cell is allocated one specific band in the frequency spectrum.
- To explore a frequency reuse pattern, the frequency spectrum is divided among seven cells, improving the voice quality as each subscriber is given a larger bandwidth.
- AMPS and ETACS cellular radio systems deploy cell-sites with tall towers that support several receiving antennas and have transmitting antennas that typically radiate a few hundred watts of effective radiated power.
- Each cell-site has one control channel transmitter that broadcasts on the forward control channel, one control channel receiver that listens on the reverse control channel for any mobile phone to set-up a call, and eight or more FM duplex voice channels.

Table 1.1 shows the worldwide 1G analog cellular system.

- All these systems use two separate frequency bands for forward (from cell-site to mobile) and reverse (from mobile to cell-site) links.

- Such a system is referred to as a frequency division duplex (DD) scheme.
- The typical allocated overall band in each direction, for example, for AMPS, and NMT-900, is 25 MHz in each direction.

- The dominant spectra of operation for these systems are the 800-and 900-MHz bands. In an ideal situation, all countries should use the same standard and the same frequency bands.
- However, in practice, as shown in Table 1.1, a variety of frequencies and standards are adopted all over the world

Table 1.1 *Existing 1G analog cellular systems*

Standard	*orward frequency band (MHz)*	*Reverse frequency band (MHz)*	*Channel spacing (kHz)*	*Number of channels*	*Multiple access/ Modulation technique*	*Major region of operation*
AMPS	824–849	869–894	30	832	FDMA/FM	US
ETACS	872–905	917–950	25	1240	FDMA/FM	UK
NMT 900	890–915	935–960	12.5	1999	FDMA/FM	EU
JTACS/ NTACS	915–925 898–901 918.5–922	860–870 843–846 863.5–867	25/12.5 25/12.5 12.5	400/800 120/240 280	FDMA/FM	Japan

AMPS Advanced Mobile Phone System
ETACS Enhanced Total Access Communication System
NMT Nordic Mobile Telephone
JTACS Japanese Total Access Communication System; NTACS: Narrowband JTACS

- All the 1G cellular systems use analog frequency modulation (FM) for which the transmission power requirement depends on the transmission bandwidth.
- power is also related to the coverage and size of the cells.
- Therefore, one can compensate for the reduction in transmission bandwidth per subscriber by reducing the size of a cell in a cellular network.
- Reduction in size of the cell increases the number of cells and the cost of installation of the infrastructure. The channel spacing, or bandwidth, allocated to each subscriber is either 30 kHz or 25 kHz or a fraction of either of them.

- **Q.2 Explain the 2G Analog Cellular system**

1.4.2 Second-Generation Digital Cellular Systems

1) The second generation (2G) cellular systems represent the set of wireless air interface standards that rely on digital modulation and sophisticated digital signal processing in the handset and the base station.
2) Digital cellular technologies support a much larger number of mobile subscribers within a given frequency allocation, thereby offering higher user capacity, providing superior security and voice quality, and lay the foundation for value-added services (including data) that will continue to be developed and enhanced in future.
3) To have efficient use of the frequency spectrum, time division or code-division multiple access technique is used in 2G digital cellular systems so that low-rate data along with voice can be processed.

There are four major standards in this category:

The North American Interim Standard (IS-54) that later on improved into

- IS-136;

- GSM,
- the pan-European digital cellular; and
- Personal digital cellular (PDC)

- All of them using TDMA technology; and IS-95 in North America, which uses CDMA technology.

- The 2G digital cellular systems are all FDD and mostly operate in the 800- and 900-MHz bands.

- The carrier spacing of IS-54/136 and PDC is the same as the carrier spacing of 1G analog cellular system in their respective regions, but GSM and IS-95 use multiple analog channels to form one digital carrier.

- The most popular 2G cellular standards include three TDMA standards and one CDMA standard. Interim

- Standard 54 or 136 (IS-54 or IS-136), also known as US Digital Cellular (USDC), which supports three time slotted mobile subscribers for each 30-kHz radio channel in both the cellular 800 MHz and PCS 1900 MHz bands.

- Based on the analog AMPS cellular system, the TDMA system IS-54/136 was developed in the US that adds digital traffic channels. IS-54/136 uses dual-mode mobile phones and incorporates associated control channels, authentication procedures using encryption, and mobile assisted handoff.

- The IS-136 includes digital control channels which enable to provide several additional services such as identification, voice mail, SMS, call waiting, group calling, etc. The USDC systems share the same frequency spectrum, frequency reuse plan, and cell-sites as that of AMPS.

- Global System for Mobile (GSM), which supports eight time slotted mobile subscribers for each 200-kHz radio channel in both the cellular and PCS bands; and Pacific Digital Cellular (PDC),

- A Japanese TDMA standard that is similar to IS-136, are the other two most popular TDMA based digital cellular standards.

- The popular 2G CDMA standard (IS-95), also known as cdmaOne, can support up to 64 mobile subscribers that are orthogonally coded and simultaneously transmitted on each 1.25 MHz channel.

- In digital communications, information is transmitted in packets or frames.
- The duration of a packet/frame in the air should be short enough, so that the channel does not change significantly during the transmission, and long enough, so that the required time interval between packets is much smaller than the length of the packet.
- A frame length of around 5 to 40 ms is typically used in 2G cellular networks..

1.4.3 Evolution from 2G to 3G Cellular Networks

There are two steps of 3G evolution paths from present 2G technologies based on GSM and IS-95 CDMA respectively. An evolution path from second generation digital cellular GSM network to third generation network is depicted in Fig. 1.1.

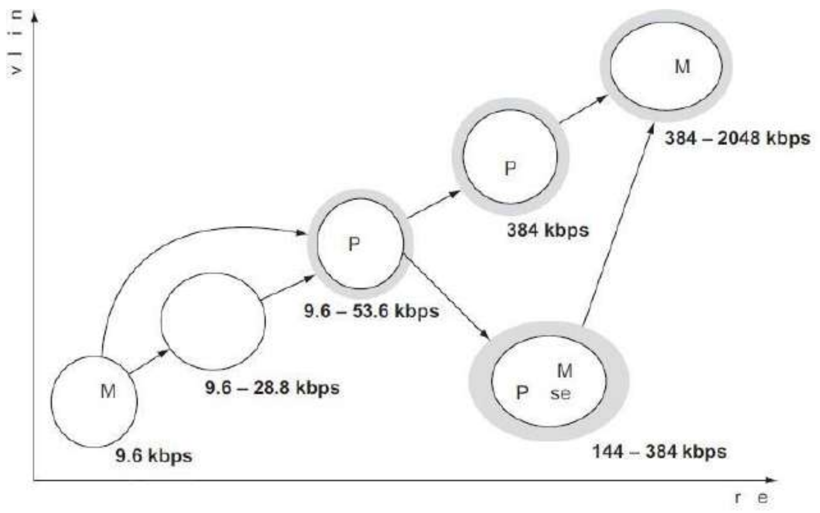


Fig. 1.1 | *An evolution path from GSM to 3G network*

- **GSM** is an open, digital cellular technology which supports voice calls and data transfer speeds of up to 9.6 kbps, together with the transmission of SMS (Short Message Service).

- GSM operates in the 900 MHz and 1.8 GHz bands in Europe and the 850 MHz and 1.9 GHz bands in the US.

- GSM provides international roaming capability that enables users to access the seamless services when travelling abroad. HSCSD (High Speed Circuit Switched Data) enables data to be transferred more rapidly than the standard

- GSM system by using multiple channels. GPRS is a very widely deployed wireless data service, available now with most GSM networks. GPRS offers throughput rates of up to 53.6 kbps

- Enhancements to GSM networks are provided by Enhanced Data rates for GSM Evolution **(EDGE)** technology or EGPRS, which offers up to three times the data capacity of GPRS.
  - Various mobile data services such as multimedia messaging, high-speed Internet access and e-mail are possible
  - **EDGE** allows it to be overlaid directly onto an existing GSM network with simple software- upgrade.

- **WCDMA i**s the air interface for third-generation mobile communications systems. It enables the continued support of voice, text and MMS services in addition to richer mobile multimedia services..

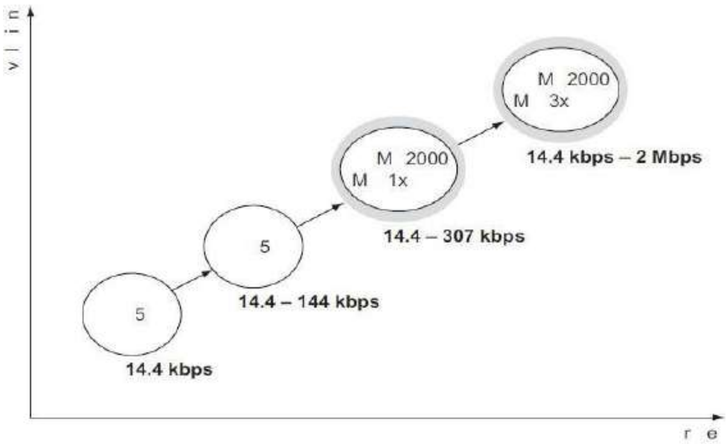


Fig.1.2 *An evolution path from CDMA to 3G network*

- **Q.3 Explain the 3G Analog Cellular system**

1.4.4 Third-Generation Digital Cellular Systems

1) The fundamental purpose of the 3G mobile communications system is to provide a globally integrated wireless communication system combining different incompatible network technologies already deployed across the world.
2) All 2G and 2.5G cellular communications systems and mobile phones will eventually evolve towards a global standard, which is referred to I**MT-2000.**

3) The third generation aims to combine telephony, Internet, and multimedia into a single device. it supports the Internet protocols and be based on a packet-switched network backbone.

The IMT-2000 system incorporates three variations of CDMA.

- The modes differ in how duplexing is accomplished and how many carriers are used.
- All variations operate in a 5-MHz channel, as compared to 1.25 MHz for cdma One systems.

Q.4 List out the 3G cellular network capabilities

(a) High degree of worldwide commonality of design
(b) Compatibility of services with fixed networks and within IMT-2000
(c) More efficient use of the available spectrum
(d) Voice quality comparable to that of PSTN
(e) 144–kbps data rate available to users in high-speed vehicles over large areas
(f) 384 kbps available to pedestrians standing or moving slowly over small areas
(g) Support for 2-Mbps data rate for office use
(h) Symmetrical and asymmetrical data-transmission rates
(i) Support for both circuit-switched and packet-switched data services
(j) Support for wide variety of mobile phones for worldwide use including pico, micro, macro, and global cellular/satellite cells
(k) Worldwide roaming capability
(l) Capability for multimedia applications and a wide range of services
(i) Flexibility to allow the introduction of new services and technologies

Q.5 List and Explain the services offered by 3G Cellular Networks

In order of increasing data rate requirements, the services are the following:

1) **Voice 3G systems** will offer speech quality at least as good as the fixed telephone network. Voicemail will also be eventually integrated fully with email service through computerized voice recognition and synthesis techniques.

2) **Switched data** This includes dial-up access to corporate networks or fax service or the Internet access that doesn't support a fully packet-switched network.

3) **Messaging** This is an extension of paging, combined with Internet e-mail service. Unlike the text-only messaging services built into some 2G systems, 3G systems will allow e-mail attachments. It can also be used for payment and electronic ticketing.

4) **Multimedia** Messaging Service (MMS) The MMS is designed to allow rich text, colour, icons and logos, sound clips, photographs, animated graphics, and video clips. It works over the broadband wireless channels in 3G networks.

5) **Immediate messaging** MMS features push capability that enables the message to be delivered instantly if the called mobile user is active. It avoids the need for collection from the server. This always-on characteristic of the mobile users opens up the exciting possibility of multimedia chat in real time.

6) **Medium multimedia** This is likely to be the most popular 3G service. Its downstream data rate is ideal for web surfing, games, location-based maps, and collaborative group working.

7) **High multimedia** This can be used for very high-speed Internet access, as well as for high-definition video and CD-quality audio on demand. Another possible application is online shopping for intangible products that can be delivered over the air such as a software program for a mobile computer.

8) **Interactive high multimedia** This can be used for high-quality videophones, videoconferencing or a combination of videoconferencing and collaborative working.

9) **Sending multimedia postcards** A clip of a holiday video could be captured through the integral video cam of a user's mobile handset or uploaded via Bluetooth from a standard camcorder, then combined with voice or text messages and mailed instantly to any other mobile user.

1G	Used analogue technologies: AMPS, NMT, TACS, J-TACS and C-Netz.
2 G	GSM, D-AMPS and IS-95
3- G	CDMA2000 and UMTS
4-G	LTE for fourth
5- G	NR

Q.6 Compare 1G, 2G, 3G, 4G and 5 G wireless communications

Technologies / Features	*1G*	*2G/2.5G*	*3G*	*4G*	*5G*
Evolution	1970	1980	1990	2000	2010
Deployment	1984	1999	2002	2010	2015
Data Rate	2 kbps	14.4-64 kbps	2 Mbps	200 Mbps to 1 Gbps for low mobility	10 Gbps to 100 Gbps
Famous Standards	AMPS	2G: GSM,C]DMA 2.5G: GPRS, EDGE, 1xRTT	WCDMA, CDMA-2000	LTA, WiMAX	Not yet defined
Technology behind	Analog cellular technology	Digital cellular technology	Broad bandwidth CDMA, IP technology	Undefined IP and seamless combination of broadband. LAN/WAN/PAN/ WLAN	Undefined IP and seamless combination of broadband. LAN/WAN/ PAN/WLAN
Service	Voice	2G: Digital Voice, SMS 2.5G: Voice+Data	Integrated high quality audio, video and data	Dynamic information access, wearable devices	Dynamic information access, wearable devices with AI capabilities
Multiplexing	FDMA	TDMA,CDMA	CDMA	CDMA	CDMA
Type of Switching	Circuit	2G: Circuit 2.5G: Circuit and packet	Packet	Packet	Packet
Handoff	Horizontal	Horizontal	Horizontal	Horizontal and Vertical	Horizontal and Vertical
Core Network	PSTN	PSTN	Packet network	Internet	Internet

1.4.5 Wireless Networking Technologies

1) In 1971 at the University of Hawaii to create the first packet-based radio communications network called ALOHAnet,

2) The very first wireless local area network (WLAN). It consisted of 7 computers that communicated in a bi-directional star topology.
- The first generation of WLAN technology used an unlicensed ISM band of 902–928 MHz.

3) To minimise the interference from small appliances and industrial machinery, a spread spectrum was used which operated at a 500-kbps data rate.

4) In 1990, the IEEE 802 Executive Committee established the 802.11 Working Group to create the WLAN standard.
5) The standard specified an operating frequency in the 2.4-GHz ISM band. In 1997, the group approved IEEE 802.11 as the world's first WLAN standard with data rates of 1 and 2 Mbps. Like cellphones, wireless-equipped laptops within range of a given access point have the ability to communicate with the network.

6) A single access point can communicate with multiple wireless equipped laptops. Many systems allow roaming between access points. Despite their limited range (up to 100 m) and lower data rates (as compared to 1 Gbps offered by wired Ethernets),

7) WLANs have become the preferred Internet access method for e-mail and Web browsing applications, in many offices, homes,campus environments, and public places.

8) A wireless personal area network (WPAN), such as Bluetooth IEEE 802.15.1, enables wireless communication between devices, ranging from computers and cell phones to keyboards and headphones, and operates in ISM 2.4 GHz band. WiMAX (WMAN based on the IEEE 802.16 family of standards) will soon offer wireless broadband Internet access to residences and businesses at relatively low cost.

2.4 BASIC PROPAGATION MECHANISMS

Q.7 In brief Explain the basic signal propogation Mechanisms

In a wireless signal propagation environment, apart from direct waves, the receiver will get a number of reflected waves, diffracted waves and scattered waves.

A typical propagation effect in a mobile radio environment is illustrated in Fig. 2.1.

As shown in Fig.2.1 ht is the height of the cell-site antenna from the earth's surface, hr is the height of the mobile antenna from the earth's surface, and r is the distance between the cell-site and the mobile unit.

The three basic propagation mechanisms are reflection, diffraction, and scattering which influence signal propagation in a mobile communication environment are briefly described now.

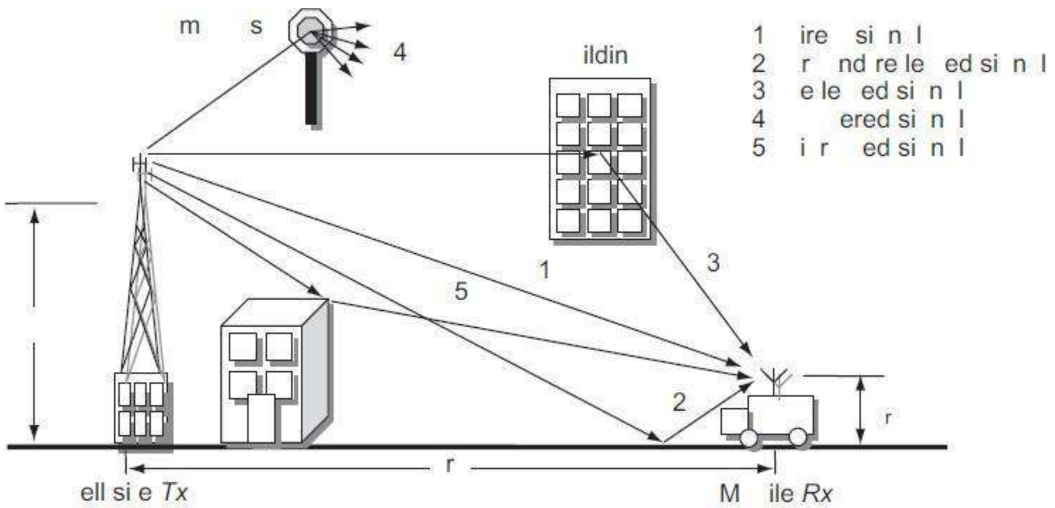


Fig. 2.1 | *Radio propagation mechanisms in a mobile communication*

2.4.1 Reflection

- Reflection occurs when incident electromagnetic waves are partially reflected when they impinge on obstructions of different electrical properties.
- A propagating electromagnetic wave impinges on objects the sizes of which are large compared to its wavelength, such as the surface of the earth, buildings, walls, etc.
- The electromagnetic radio waves get reflected from tall building structures which have a good amount of conductivity.
- Reflection can also occur due to metal reinforcement.
- The extent of reflection of radio waves depends on the composition and surface characteristics of the objects.
- The angle of reflection is equal to the angle at which the wave strikes the object and is measured by the Fresnel reflection coefficient.
- Upon reflection, the signal strength of the radio wave gets attenuated that depends on many factors like the frequency of the radio waves, the angle of incidence, and the nature of the medium including its material properties, thickness, homogeneity, etc. Generally, higher frequencies reflect more than lower frequencies.
- As an instance, let a ground-reflected wave near the mobile unit be received. Because the ground-reflected wave has a 180o phase shift after reflection, the ground wave and the line-of-sight wave may tend to cancel each other, resulting in high signal attenuation.
- The vector sum of the phases of the multipath received signals may give a resultant zero amplitude at certain time instants and large signal amplitude at some other time.
- Most of the times, the vectorial addition of these multipath reflected signals produce an undetectable signal. Further, because the mobile antenna is lower than most human-made structures in the operational area, multipath interference occurs.
- These reflected waves may interfere constructively or destructively at the receiver.
- In outdoor urban areas, the reflection mechanism often loses its importance because it involves multiple reflections that reduce the strength of the signal to negligible values. However, reflection mechanisms often dominate radio propagation in indoor applications.
- The reflections are a source of multipath signals which cause low strength in signal reception. Reflection results in a large-scale fading of the radio signals.

EXAMPLE 2.3 Effects of reflection on signal propagation

A wireless communication transmitter transmits a signal at 900 MHz. A receiver located at a distance of 1 km away from transmitter receives two signals — one directly as a line-of-sight signal and another indirectly via reflection from a building having a height more than 10 metres), as shown in Fig. 2.2.

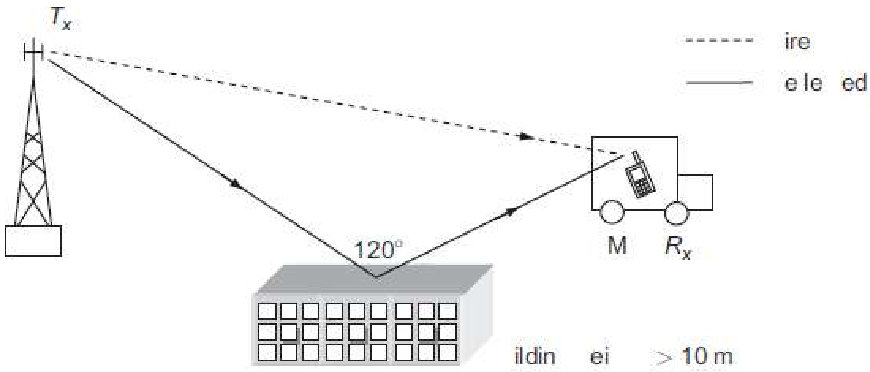


Fig. 2.2 | *Reflected and direct signals*

Problem 1

Give reason(s) to justify that the reflected signal causes delay in the reception. Calculate the amount of delay in the reflected signal with respect to the direct signal at the receiver

Solution

Frequency of transmission, $f_c = 900$ MHz (given)

Step 1. *To find the wavelength of transmission, λ_c*

We know that $\lambda_c = c / f_c$

Or, $\lambda_c = 3 \times 10^8$ m/s / 900×10^6 Hz

Therefore, $\lambda_c = 0.33$ m

Step 2. *To justify that reflected signal causes delay.*

The height of the building 10 m (given)

Thus, the given height of the building is much greater than the wavelength of the transmission. It implies that the radio signal is reflected from the surface of the obstacle of size much greater than λc of the radio transmissions. The reflected signal suffers a delay in reaching the receiver

Step 3. *To find the time taken by the direct path, t_{direct}*

Distance between transmitter and receiver = 1 km or 1000 m (given)

We know that t_{direct} = distance traveled by direct path / speed of radio wave

Or, $t_{direct} = 1000$ m / 3×10^8 m/s

Therefore, $t_{direct} = 3.33$ μs

Step 4. *To find the time taken by the reflected path, $t_{reflected}$*

Assuming that the reflected path is approximately equal to 1000 m

Angle between incident and reflected path = 120 (given)

Thus, incident angle = 120 /2 = 60

Therefore, $t_{reflected} = 1000$ m / $(3 \times 10^8$ m/s$) \times \sin 60$

Or, $t_{reflected} = 3.85$ μs

Step 5. *To calculate the delay in a reflected signal*

Delay = $t_{reflected}$ t_{direct}

Hence, delay = 3.85 μs 3.33 μs = 0.52 s

2.4.2 Diffraction

- Diffraction is referred to the change in wave pattern caused by interference between waves that have been reflected from a surface or a point.

- It is based on Huygen's principle which states that all points on a wavefront can be considered as point sources for production of secondary wavelets that can combine to produce a new wavefront in the direction of propagation of the signal.

- Diffraction occurs when the radio path between a transmitter and receiver is obstructed by a surface with sharp irregular edges.

- Waves bend around the obstacle, even when a line-of-sight condition does not exist. It causes regions of signal strengthening and weakening irregularly.

- Diffraction can also occur in different situations such as when radio waves pass through a narrow slit or the edge of a reflector or reflect off from two different surfaces approximately one wavelength apart.

- At higher frequencies, diffraction depends on the geometry of the object, as well as the amplitude, phase, and polarization of the incident wave at the point of diffraction. Figure 2.3 depicts a simple case of diffraction of a radio signal.

- Diffraction is a description of how a radio signal propagates around and over an obstruction, and is measured in dB. Diffraction often results in small-signal fading.

- In effect, diffraction results in propagation into shadow regions because the diffracted field can reach a receiver, which is not in the line-of-sight of the transmitter.

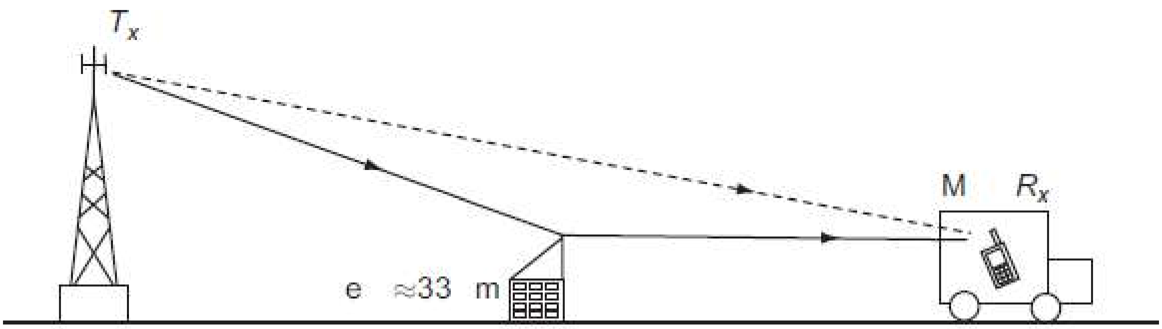


Fig. 2.3 | *Diffraction of a radio signal*

- In mobile communication systems, diffraction loss occurs from the blockage of secondary waves such that only a portion of the energy is diffracted around an obstacle. Most cellular systems operate in urban areas where there is no direct line-of-sight path between the transmitter and the receiver (either from the cell-site to the mobile unit or vice-versa), and where the presence of highrise buildings causes severe diffraction loss.

- In many practical situations, the propagation path may consist of more than one obstruction. For example, in hilly terrains, the total diffraction loss must be computed due to all of the obstacles.

2.4.3 Scattering

- Scattering is a special case of reflection caused by irregular objects such as walls with rough surfaces, vehicles, foliage, traffic signs, lamp posts, and results in many different angles of reflection and scatter waves in all directions in the form of spherical waves. Thus, due to availability of numerous objects, scattering effects are difficult to predict.

- Scattering occurs when the size of objects is comparable or smaller than the wavelength of the propagating radio wave, and where the number of obstacles per unit volume is large. Figure 2.4 depicts a typical case of scattering of a radio signal.

- Propagation in many directions results in reduced received-signal power levels, especially far from the scatterer. So an incoming radio signal is scattered into several weaker outgoing radio signals.

- As a result, the scattering phenomenon is not significant unless the receiver or transmitter is located in a highly noisy environment. In a mobile radio environment, scattering provides additional radio energy levels at the receiver to what has been predicted by reflection and diffraction models alone. In radio channels,

- Knowledge of the physical location of large distant objects, which induce scattering, can be used to accurately predict scattered signal strength levels.
- In a mobile radio environment, heavy foliage often causes scattering.
- Scattering too results in small-scale fading effects.

Major adverse effect of multipath propagation

1) Multiple copies of a signal may arrive at different phases. If these phases add destructively, the signal level relative to noise declines, making signal detection at the receiver much more difficult and unreliable.

2) Increase in received data errors due to intersymbol interference in digital transmission. As the mobile unit moves, the relative location of various objects also changes; hence intersymbol interference increases to the extent that makes it difficult to design signal processing techniques that will filter out multipath effects in order to recover the intended signal with fidelity.

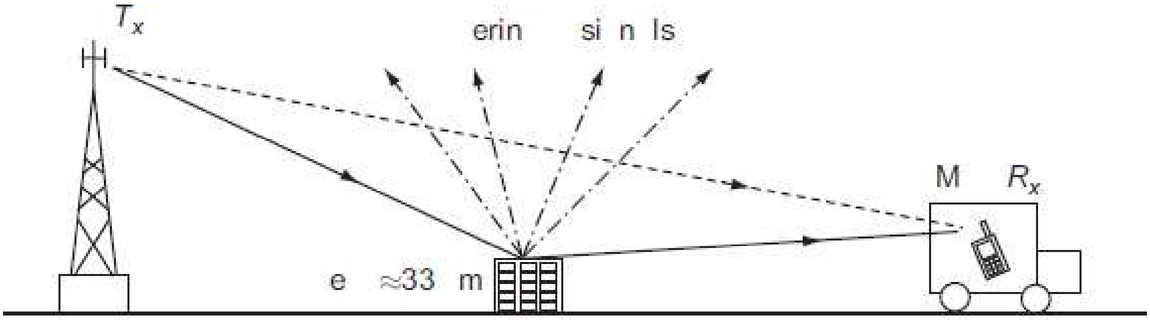


Fig. 2.4 | *Scattering of radio signal*

Refraction

- Another form of propagation effect is the effect of refraction.
- Refraction occurs because the velocity of the electromagnetic waves depends on the density of the medium through which it travels.
- Waves that travel into a denser medium are bent towards the medium. This is the reason for line-of-sight radio waves being bent towards the earth since the density of the atmosphere is higher closer to the earth

2.5 MOBILE RADIO CHANNEL

Q.8 what do you mean by Fading and Multipath fading

- In a mobile communication system, a signal experiences multipath propagation which causes rapid signal level fluctuations in time, called fading.
- Mobile radio channels introduce noise, fading, interference, and other distortions into the signals that they transmit.

Fading effects that characterise mobile radio communication are large-scale fading and small-signal fading.

Rayleigh Fading.

1) If there is a large number of multiple reflective paths with no line-of-sight signal path, it is Rayleigh fading.
2) The Rayleigh flat-fading channel model assumes that the channel induces amplitude which varies in time according to Rayleigh distribution.

Rician fading.

1) When there is a dominant non-fading signal component present, the small-signal fading envelope is described by a **Rician fading.**
2) Small-signal fading results into signal dispersion and time-variant behaviour of the channel.

Causes for Rayleigh and Rician fading phenomena

1) include multipath scattering effects, time dispersion, and Doppler shifts that arise from relative motion between the transmitter and receiver.
2) The major paths result in the arrival of delayed versions of the signal at the receiver.
3) In addition, the radio signal undergoes scattering on a local scale for each major path. Such local scattering is typically characterized by a large number of reflections by objects near the mobile.
4) These irresolvable components combine at the receiver and give rise to the phenomenon known as multipath fading. As a result, each major path behaves as a discrete fading path.

5) Typically, the fading process is characterised by a Rayleigh distribution for a non-line-of-sight path and
6) A Rician distribution for a line-of-sight path. In mobile radio channels,
7) The Rayleigh distribution is commonly used to describe the statistical time varying nature of the received envelope of a flat fading channel, or the envelope of individual multipath components.

2.5.1 Multipath Fading

- Fading of signal received by the mobile unit is an inherent problem in mobile communication.

- As the location of the mobile unit keeps on changing in real time, the resultant radio signal incident on its antenna varies continuously.
- Multipath in the mobile communication channel creates small-scale fading effects such as rapid

- changes in signal strength over a small time interval or small distance traveled by a mobile; random frequency modulation due to varying Doppler shifts on different multipath signals; and time dispersion caused by multipath propagation delays.

- Fading is the rapid fluctuation of a radio signal's amplitude in a short time or over a short distance.

- In reality, the received signal rapidly fluctuates due to the mobility of the mobile unit causing changes in multiple signal components arriving via different paths.

- These multiple waves can combine constructively or destructively.

- Multipath waves are also generated because the antenna height of the mobile unit is lower than its typical surrounding structures such as in builtup urban areas of operation, and the operating wavelength is much less than the sizes of the surrounding structures at the mobile unit.

- The sum of multipath waves causes a signal-fading phenomenon. The rapid fluctuation of the signal amplitude is referred to as small-signal fading, and it is the result of movement of the transmitter, the receiver, or objects surrounding them. Over a small area, the average value of the received signal is considered to compute the propagation path loss and received signal strength. But the characteristics of the instantaneous signal level are also important in order to design receivers that can mitigate these effects.

- Multipath fading results in fluctuations of the signal amplitude because of the addition of signals arriving with different phases. This phase difference is caused due to the fact that signals have traveled different path lengths.
- Because the phase of the arriving paths are too changing rapidly, the received signal amplitude undergoes rapid fluctuation that is often modeled as a random variable with a particular distribution, called Rayleigh distribution.
- The multipath waves at the mobile receiver bounce back and forth due to the surrounding buildings and other structures, as shown in Fig. 2.5. When a mobile unit is stand-still, its receiver only receives a signal strength at that spot, so a constant signal is observed. When the mobile unit is moving, the fading structure of the wave in the space is received. It is a multipath fading which becomes fast as the vehicle moves faster.

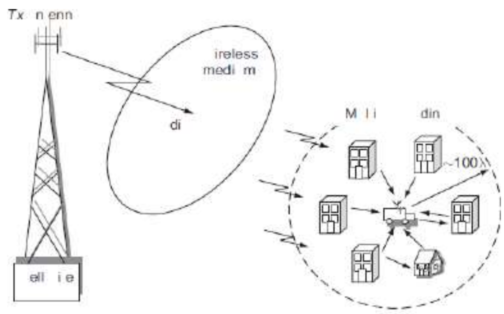


Fig. 2.5 | *Multipath fading in a mobile radio environment*

Reasons that contribute to the rapid fluctuations of the signal amplitude

1) The first, caused by the addition of signals arriving via different paths, is referred to as multipath fading.
2) The second, caused by the relative movement of the mobile unit towards or away from the cell-site transmitter, is called Doppler effect. Other factors that influence small-scale fading include multipath propagation, speed of the mobile, speed of the surrounding objects, and the transmission bandwidth of the signal.
3) For a particular service area, the fading effects of the received signal at the mobile unit need to be analysed towards the effort of designing a reliable mobile communication system. Suitable diversity reception or signal-processing techniques need to be provided to minimise the impact of fading.

Q,9 List and Eplain the types of small scale Fadings

2.5.2 Types of Small-Scale Fading

1) The type of fading experienced by a signal propagating through a mobile communication channel depends

- On the nature of the transmitted signal with respect to the characteristics of the wireless channel,
- The speed of the mobile, and
- The direction of motion of the mobile with respect to the incoming received signal from the cell-site transmitter.
- Fading effects in a mobile radio environment can be classified as
  – fading effects due to multipath time delay spread; and
  – fading effects due to Doppler spread.

2) Due to multipath time-delay spread, fading effects can also be classified as

- Flat fading :
- Frequency selective fading.

2.5.3 Flat fading, or non-selective fading, is that type of fading in which all frequency components of the received signal fluctuates in the same proportions simultaneously.

- Flat fading occurs when the radio channel has a constant gain and linear phase response but its bandwidth is greater than that of the transmitted signal.

- It implies that the desired signal bandwidth is narrower than, and completely covered by, the spectrum affected by the fading.

- In flat fading, the multipath structure of the channel is such that the spectral characteristics of the transmitted signal are preserved at the receiver. However, the strength of the received signal changes with time due to fluctuations in the gain of the channel caused by multipath.

- In a flat fading channel, sometimes referred to as a narrowband channel, the bandwidth of the transmitted signal is much larger than the reciprocal of the multipath time-delay spread of the channel.

- The bandwidth of the applied signal is narrow as compared to that of the wireless channel. The distribution of the instantaneous gain of flat-fading channels can be best described by Rayleigh distribution and is important for designing wireless communication links.

Frequency-selective fading affects unequally the different spectral components of a radio signal.

- Selective fading is usually significant only relative to the bandwidth of the overall wireless communication channel.

- If the signal attenuates over a portion of the bandwidth of the signal, the fading is considered to be selective in frequency domain.

- Frequency selective fading on the received signal occurs when a radio channel has a constant gain and linear phase response, but the channel bandwidth is less than that of the transmitted signal.

- Under such conditions, the channel impulse response has a multipath delay spread which is greater than the reciprocal bandwidth of the transmitted signal.

  Frequency selective fading is due to time dispersion of the transmitted symbols within the channel, and the channel induces intersymbol interference.

  Frequency-selective fading channels are also known as wideband channels since the bandwidth of the transmitted signal is wider than the bandwidth of the channel impulse response.

As an example, suppose a mobile receiver moves directly away from the transmitting antenna but toward a reflecting surface. This particular scenario is depicted in Fig. 2.6.

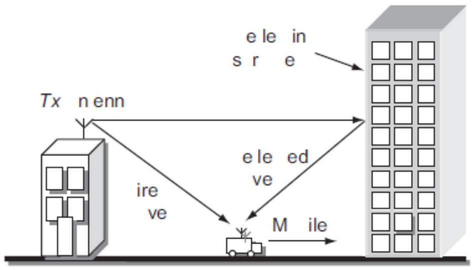


Fig. 2.6 | *Fast fading in a mobile environment*

Q,10 Explain the multipath delay spread

2.5.4 Multipath Delay Spread

- Multipath interference is the reflection of radio signals from concrete structures that results in multiple copies of the received signal.
- Multipath interference can allow radio signals to reach hard-to-reach areas.
- It can also create some problems such as delay spread which occurs when several signals reach a receiver at different times due to different lengths of transmission paths. Delay spread also occurs due to Rayleigh fading which results from the signal's amplitude and phase being altered by reflections.
- In a digital communication system, the delay spread along with fading causes intersymbol interference, thereby limiting the maximum symbol rate of a digital multipath channel.
- If the multipath delay spread is comparable to or larger than the symbol duration, the received waveform spreads into neighbouring symbols and produces intersymbol interference. The intersymbol interference results in irreducible errors that are caused in the detected signal.

- Figure 2.8 shows the multiple signals received at different multipaths.

2.5.5 Doppler Shift

- There is always a relative motion between the cell-site transmitter and the mobile receiver.
- As a result, Doppler effect occurs in the shift of the received carrier frequency.
- Doppler spectrum is the spectrum of the fluctuations of the received signal strength. Multipath fading provides the distributions of the amplitude of a radio signal.
- It is important to know for what time a signal strength will be below a pre-defined threshold value, that is, the duration of fade, and how often it crosses a threshold value, that is, frequency of transitions or fading rate.
- Doppler effect results in the inaccurate operation of the system. Proper compensation technique needs to be implemented to minimise this effect.
- A study of Doppler spectrum is important to design the coding and interleaver schemes for efficient performance.

Thus, multipath propagation, speed of mobile unit, speed of reflecting objects, and Doppler shift are the main causes of fading.

2.5.6 Coherence Bandwidth

- The coherence bandwidth is a statistical measure of the range of frequencies over which the channel can be considered flat.
- A flat channel is one which passes all spectral components with approximately equal gain and linear phase and without any distortion.
- The coherence bandwidth Bc represents the correlation between two fading signal envelopes at frequencies f1 and f2 and is a function of the delay spread d.
- When the correlation coefficient between two fading signal envelopes at frequencies f1 and f2 is equal to 0.5, the coherence bandwidth Bc is approximated by:

$$Bc \approx 1 / (2 \pi d) \quad (2.15)$$

- where d is the delay spread.

2.5.7 Coherence Time

- Coherence time is the time duration over which two received signals have a strong potential for amplitude correlation.
- In other words, coherence time c is inversely proportional to the Doppler spread.
- It is used to characterise the time-varying nature of the frequency dispersiveness of the channel in the time domain.
- If the reciprocal bandwidth of the baseband signal is greater than the coherence time of the channel then the channel will change during the transmission of the baseband signal, thus causing distortion at the receiver.
- If the coherence time is defined as the time over which the time correlation function is above 0.5, then the coherence time is approximately given by

$$c \approx 0.423 / fdm \quad (2.18)$$

- where fdm is the maximum Doppler shift given by Vm / λc.

Principles of Cellular Communication

1) Cellular communication is designed to enhance the spectrum efficiency as well as the system capacity while maintaining the desired signal quality.
2) The main principle of cellular communication is to divide a large geographical area into a number of contiguous smaller geographical coverage areas called cells, each one of which is served by its own cell-site or low-power base station located at its centre.
3) Cells constitute the design of the heart of cellular systems.
   - The focus in this chapter is to understand the essential principles of cellular communication, and
   - The formation of regular hexagonal cellular structures with multiple clusters.
   - The most serious concern due to frequency reuse is cochannel interference which may degrade the performance of a cellular system operation. Finally, a brief overview of various methods employed to reduce cochannel interference is also given.

Q.11 Explain the cellular Terminologies

4.1 CELLULAR TERMINOLOGY

- A cell is the basic geographic unit of a cellular system.
- A cell is the radio area covered by a cell-site that is located at its centre. In other words, the radio coverage by one base station or a cell-site is referred to as a cell, which is also called a footprint.
- In a cellular system, the most important factor is the size and shape of a cell. Because of constraints imposed by natural irregular terrain, man-made structures, and non uniform population densities, the actual shape of the cell may not be either a circle or a regular geometrical shape but may be a little distorted.
- For proper analysis and evaluation of a cellular system, an appropriate model of a cell shape is needed. Figure 4.1 depicts ideal cell, actual cell and possible cell models such as equilateral triangle, square, and hexagon that represent a cell boundary with a radius R from the centre of the cell.

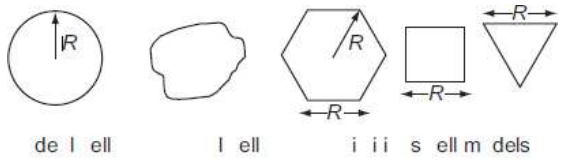


Fig. 4.1 | *Ideal, actual and fictitious cell models*

- The actual shape of the cell is determined by the desired received signal level by the mobile subscribers from its base-station transmitter in its operating area.

- The received signal is affected by many factors including reflections, refractions, and contour of the terrain as well as multipath propagation due to presence of natural and man-made structures.
- A cell is not a perfect polygon. So real footprints are vague in nature.
- On the other hand, cellular layouts using irregular structures limit growth and are also inefficient. For this reason, cellular layouts and performance studies are based on regular topologies as they allow the systematic growth though they may be just conceptual.
- The base station, also called Cell-Site (CS), located approximately at its centre, serves all mobile users in the cell.

- Figure 4.2 illustrates an ideal cell area (circular), a hexagonal cell area (used in most models), and a square cell area (an alternative shape) with a cell-site at its centre and a number of mobile units (M) within the cell area.
- The shape of the cell can be circular around the cell-site transmitting tower under ideal radio environment.
- The periphery of the circle is equal to the acceptable received signal level from the transmitting signal. It means that if the cell-site is located at the centre of the cell, the cell area and periphery are determined by the signal strength within the region.
- This depends on many factors, such as the height of the cell-site transmitting antenna; contour of the terrain; presence of tall buildings, hills, valleys, vegetation; and atmospheric conditions.
- Therefore, the actual shape of the cell may be a zigzag shape which indicates a true radio coverage area. However, for all practical purposes, a regular hexagonal geometry shape approximates the cell boundary, which is a good approximation of a circular region. However, the square is another alternative shape that can be used to represent the cell area.

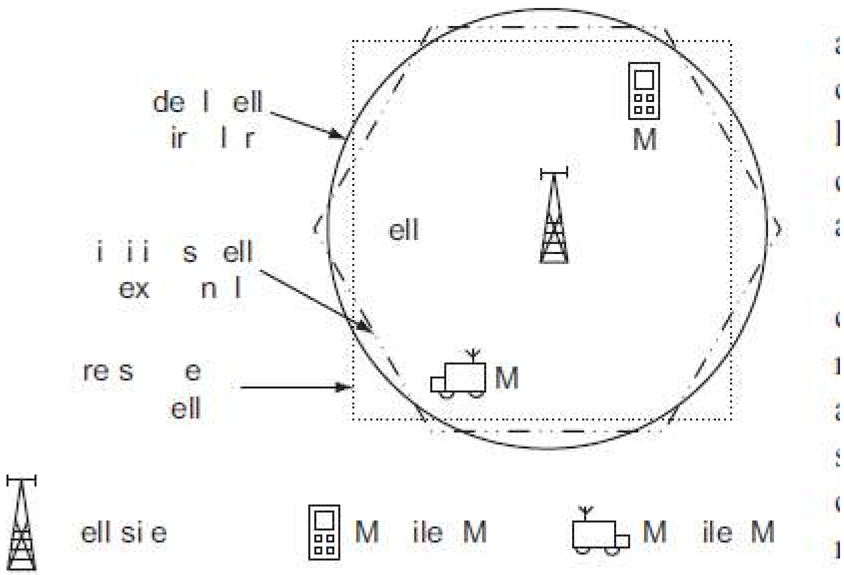


Fig. 4.2 | *Illustration of a cell with a CS and Mobile*

Q,12 Explain the cell structure and Cluster

4.2 CELL STRUCTURE AND CLUSTER

1) In practice, cells are of arbitrary shape which is quite close to a circle, is the ideal radiation pattern of an omnidirectional antenna. Because of the randomness inherent nature of the mobile radio propagation and irregular geographical terrain,
2) it is easier to obtain insight and plan the cellular network by visualising all the cells as having the same shape.
3) By approximating a uniform cell size for all cells, it is easier to analyse and design a cellular topology mathematically. It is highly desirable to construct the cellular system such that the cells do not overlap, and are tightly packed without any dead signal spots.
4) The cellular topology formed by using ideal circular shape results into overlaps or gaps between them which is not desirable in cellular communications which has to be essentially continuous.

5) This form of layout requires the use of regular topologies (say, a hexagonal topology) instead of a circular shape, as depicted in Fig. 4.3.

In Fig. 4.3, the middle dark circles represent cell-sites. This is where the base-station radio equipment and their antennas mounted on tall towers are located.

- A cell-site gives radio signal coverage to a cell. In other words, the cell-site is a location or a point at the centre of the cell, whereas the cell is a wide geographical service area.

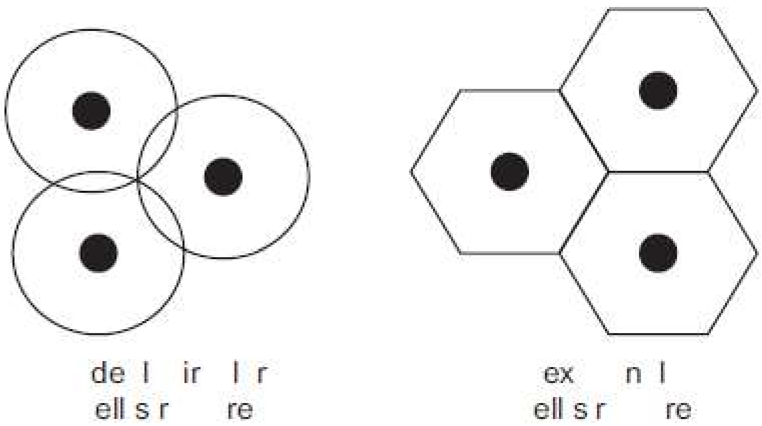


Fig. 4.3 *The ideal and regular hexagonal cell structure*

- The design and performance of cellular systems using regular geometrical topologies may not correspond to real mobile environments, but these topologies do provide valuable information and guidelines for structuring practical cellular configuration layouts.
- Cells of the same shape form a tessellation so that there are no ambiguous areas that belong to multiple cells or to no cell.
- The cell shape can be of only three types of regular polygons: equilateral triangle, square, or regular hexagon as shown in Fig. 4.4.

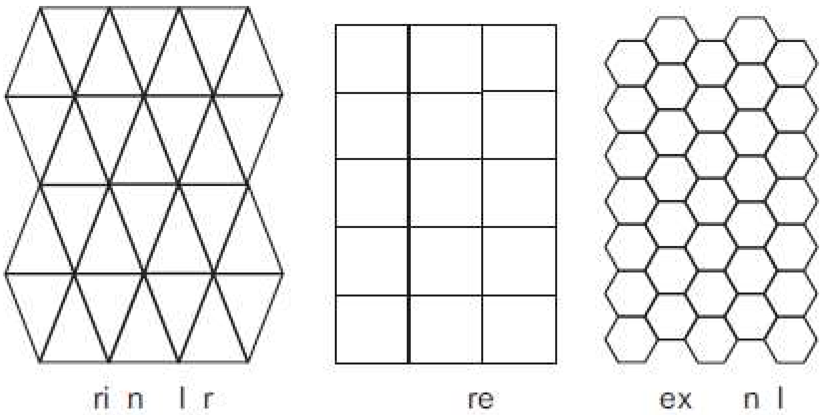


Fig. 4.4 *Illustration of possible geometrical cellular structures*

- A cellular structure based on a regular hexagonal topology, though fictitious, offers best possible non-overlapped cell radio coverage.
- Traditionally, a regular hexagonal-shaped cell is the closest approximation to a circle out of these three geometrical shapes and has been used for cellular system design. In other words,

for a given radius (largest possible distance between the polygon centre and its edge), the hexagon has the largest area. Moreover, it allows a larger region to be divided into nonoverlapping hexagonal subregions of equal size, with each one representing a cell area.

- Octagons and decagons geometrical patterns do represent shapes closer to a circular area as compared to hexagons, but they are not used to model a cell as it is not possible to divide a larger area into non- overlapping subareas of the same size.
- A mobile radio communication system is generally required to operate over areas too large to be economically covered by a single cell-site.
- Therefore, several or many widely spaced transmitter sites are required to provide total area coverage.
- The spacing between the base stations need not be regular and the cell or the area served by a base station need not have any particular shape. However, the absence of an orderly geometrical structure makes the system design more difficult and results in inefficient use of spectrum and uneconomical deployment of equipment.
- The propagation considerations recommend the circle as a cell shape for defining the area covered bya particular base station. This is impracticable for design purpose, since there could be areas which are contained either in no cell or in multiple cells. On the other hand, any regular polygon can cover the service area with no gaps or overlaps.
- The regular hexagonal shape results in the most economical system layout design.

In most modeling, simulation, measurements, and analysis of interference in cellular systems, hexagons are used to represent the cell structure.

- A hexagon is closer to a circular area and multiple hexagons can be arranged next to each other, without having an overlapping area or uncovered space in between. In other words, the hexagonal-shaped cells fit the planned area nicely, with no gap and no overlap among the adjacent hexagonal cells. Thus, it simplifies the planning and design of a cellular system.

Problem3 important

Example 1

Consider a single high-power transmitter that can support 40 voice channels over an area of 140 km² with the available spectrum. If this area is equally divided into seven smaller areas (cells), each supported by lower power transmitters so that each cell supports 30% of the channels, then determine

(a) Coverage area of each cell

(b) Total number of voice channels available in cellular system

Comment on the results obtained.

Solution

Total service area to be covered = 140 km (given)

Total number of channels available = 40 (given)

Number of cells = 7 (given)

(a) To determine coverage area of each cell

Step 1. Coverage area of a cell = Total service area / Number of cells

Hence, coverage area of a cell = 140 km / 7 = 20 km

(b) To determine total number of voice channels available in the cellular system

Step 2. Number of voice channels per cell = 30% of original channels (given)

Number of voice channels per cell = 0.3 × 40 = 12 channels/cell

Total number of voice channels available in cellular system is given by the number of channels per cell multiplied by the number of cells in the service area.

Hence, total number of voice channels = 12 × 7 = 84 channels

1) A Cellular Cluster A group of cells that use a different set of frequencies in each cell is called a cellular cluster. Thus, a cluster is a group of cells with no reuse of channels within it. It is worth mentioning here that only a selected number of cells can form a cluster.
2) It follows certain rules before any cell can be repeated at a different location. Some common reuse cluster patterns are given in Fig. 4.5.
3) Two or more different cells can use the same set of frequencies or channels if these cells are separated in space such that the interference between cells at any given frequency is at an acceptable level. That means, the cluster can be repeated any number of times in a systematic manner in order to cover the designated large geographical service area.

Let there be K number of cells having a different set of frequencies in a cluster. Then K is termed as the cluster size in terms of the number of cells within it.

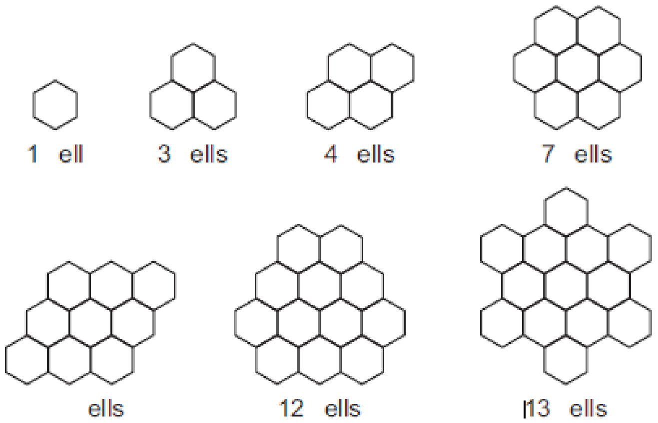


Fig. 4.5 *Common reuse patterns of hexagonal cell clusters*

problem 4

Calculate the number of times the cluster of size 4 have to be replicated in order to approximately cover the entire service area of 1765 km2 with the adequate number of uniform-sized cells of 7 km2 each.

Solution

Size of the cluster, K = 4 (given)
Area of a cell, Acell = 7 km2 (given)
Step 1. To determine area of the cluster
Area of a cluster, Acluster = K × Acell
Therefore, Acluster = 4 × 7 km2 = 28 km2
Step 2. To determine number of clusters in the service area
Total service area, Asystem = 1765 km2 (given)
Number of clusters in service area = Asystem / Acluster
Number of clusters in service area = 1765 km2 / 28 km2
Number of clusters in service area = 63
Hence, the number of times the cluster of size 4 has to be replicated is 63.

Each cell size varies depending on the landscape. Typical size of a cell may vary from a few 100 metres

- in cities (or even less at higher frequencies) to several kilometres on the countryside.

- Smaller cells are used when there is a requirement to support a large number of mobile users, in a small geographic region, or when a low transmission power may be required to reduce the effects of interference.
- So typical uses of small cells are in urban areas, low transmission power required, or higher number of mobile users.

It is clear that if the cell area is increased, the number of channels per unit area is reduced for the same number of channels and is good for less populated areas, with fewer mobile users. Generally, large cells are employed in remote areas, coastal regions, and areas with few mobile users, large areas that need to be covered with minimum number of cell-sites.
It may also be noted that the cell area and the boundary length are important parameters that affect the handoff from a cell to an adjacent cell.
A practical solution for optimum cell size is to keep the number of channels per unit area comparable to the number of mobile subscribers
to be served within that cell.

Q.13 Explain frequency Reuse concept

4.3 FREQUENCY REUSE CONCEPT

About the essence of cellular communication?

1) if a single base station serves a wireless communication system, a high power transmitter is needed to support a large number of users.
2) Moreover, due to availability of limited RF spectrum, the maximum number of simultaneous users in this system is also limited.
3) If allocated RF spectrum or a given set of frequencies (frequency channels) can be reused in a given large geographical service area without increasing the interference then the service area can be divided into a number of small areas called cells, each allocated a subset of frequencies. With smaller area coverage, lower power transmitters with lower height antennas can be used at a base station.
4) The conventional radio communication systems are faced with the problems of limited service area capability and inefficient spectrum utilisation. This is because these systems are usually designed for providing service in an autonomous geographic zone and by selecting RF channels from a specified allocated frequency band.
5) Contrary to this, the present mobile radio communication system are designed for wide area coverage and high grade of service.
6) At the same time, the systems are required to provide continuous communication through an effective usage of available spectrum. This dictates that the mobile radio network design must satisfy the objective of providing continuous and wide service area coverage while optimally using the RF spectrum.
   - The increase in system capacity is achieved with the use of smaller cells, reuse of frequencies, and cell sectoring.
   - Frequency reuse is the core concept of the cellular communications

7) In a mobile radio network designed on the basis of frequency reuse concept, it must be ensured that the service area is adequately protected from the cochannel and the adjacent-channel interference.
8) The carrierto- interference ratio (C/I) requirements are considerably lower for digital systems as compared to analog systems. It is seen that spectrum efficiency increases if the C/I value is

lowered. This is due to the fact that lowering the acceptable value of C/I reduces the frequency reuse distance and the reuse pattern. The co channel interference can be controlled by geographical separation whereas adjacent-channel interference depends on the receiver filter characteristics and out-of-band transmission.

9) The design process of selecting and allocating channel groups for all the cellular base stations within a system is called frequency reuse. Thus, large coverage area, efficient spectrum utilisation and enhanced system capacity are the major attributes of cellular communication.

10) This requires proper system design and complex operation of the cellular mobile system working in a hostile mobile propagation environment and system interference in order to ensure the desired service performance

11) A regular geometrical hexagonal pattern results in obtaining optimum area coverage and efficient spectrum utilisation.

12) The minimum value of cluster size provides optimum spectrum occupancy. However, in actual design, due to physical limitations the location of base stations cannot follow the regular geometrical hexagonal pattern. The resultant location errors distort the regular pattern, thereby causing serious interference problems.

- Mobile users communicate only via the base stations. Each cell is allocated a finite number of Radio Frequency (RF) channels, depending upon the number of simultaneous users required to be served.
- This enables the cells that are located sufficiently physically apart to reuse the same set of frequencies, without causing cochannel interference.
- However, each adjacent cell within a cluster operates on different frequencies to avoid interference. Cells, which use the same set of frequencies, are referred to as cochannel cells.
- The space between adjacent cochannel cells is filled with other cells that use different frequencies to provide frequency isolation.
- A typical cluster of seven cells, each repeated seven times with frequency reuse, is illustrated in Fig. 4.6.

<u>If the system is not properly designed, cochannel interference may occur due to the simultaneous use of the same channel. This is the major concern in frequency reuse. Specifically, if the available channels are reused for additional traffic, it is possible to</u>

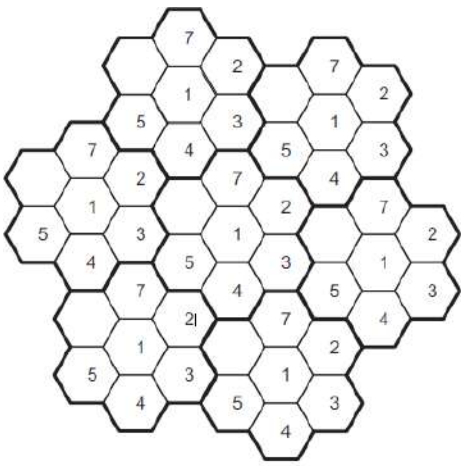


Fig. 4.6 | *Illustration of frequency reuse (Fx: A set of frequency channels)*

serve more number of users, thereby increasing the system capacity within allocated RF spectrum, and hence enhancing spectrum efficiency as well.

- The total number of channels available in a cellular system is finite because of limited RF spectrum allocation.
- The capacity of a cellular system is defined by the total number of channels available, which depends on how the available channels are deployed. So, the total number of available channels without frequency reuse, N, is the allocated RF spectrum band divided by the number of RF channels having equal channel bandwidth.

Q.14 Explain Cluster size and System capacity

4.4 CLUSTER SIZE AND SYSTEM CAPACITY

- The K number of cells in the cluster would utilise all N available channels. In this way, each cell in the cluster contains N/K number of channels only.
- Alternately, the total number of channels available in a cluster, N is equal to the number of channels per cell ($J \leq N$) multiplied by the number of cells per cluster (K), that is,

  $N = J \times K$ (4.1)
- In a cellular system, the whole geographical area where the cellular services are required to be provided is divided into a number of clusters having a finite number of cells. The K cells in a cluster use the complete set of available frequency channels.
- Since N is the total number of available channels, it can be seen that a decrease in the cluster size K is accompanied by an increase in the number of channels J allocated per cell. Thus, by decreasing the cluster size, it is possible to increase the capacity per cell.
- The cluster can be replicated many times to cover the desired geographical area by a cellular communication system. The overall system capacity, C, can then be theoretically determined by simply multiplying the number of clusters in a system (say M) with total number of channels allocated to a cluster, N, i.e.,

$C = M \times N$ (4.2)

Using the relationship $N = J \times K$, we get

$C = M \times J \times K$

- If K is decreased and J is proportionally increased so that $C = M \times J \times K$ is satisfied, it is necessary to replicate the smaller cluster more times in order to cover the same geographical service area.
- This means the value of M has to be increased. Since $J \times K$ (=N) remains constant and M is increased, it shows that the system capacity C is increased. That is, when K is minimised, C is maximised. But minimizing K will increase cochannel interference.

Q.15 Explain Frequency Reuse Distance

4.6 FREQUENCY REUSE DISTANCE

1) Reusing an identical frequency channel in different cells is limited by cochannel interference between cells and
2) The cochannel interference can become a major problem in cellular communication. So it is desirable to find the minimum frequency reuse distance D in order to reduce this cochannel interference.

The minimum distance, which allows the same frequency to be reused in cochannel cells, will depend on many factors such as

- The number of co channel cells in the vicinity of the centre cell,
- The type of geographic terrain contour,
- The antenna height, and

• The transmitted power at each cell-site.

3) Assume that the size of all the cells is approximately same; the cell size is usually determined by the coverage area of the signal strength in each cell. As long as the cell size is fixed, cochannel interference is independent of transmitted power of each cell.
4) It means that the received signal threshold level at the mobile unit is adjusted to the size of the cell.

Actually, cochannel interference is a function of a parameter known as frequency reuse ratio, q, and is defined as q = D R (4.14)

Where D is the distance between two nearest cochannel cells marked as C1 and R is the radius of the cells under consideration, as shown in Fig. 4.14.

It may be noted here that this ratio is applicable for any value of cluster size K.

5) The parameter q is also referred to as the cochannel reuse ratio or the cochannel reuse factor or cochannel interference reduction factor or frequency reuse ratio.

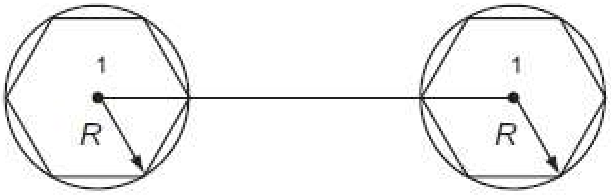


Fig. 4.14 | *Frequency reuse ratio q = D / R*

Problem 5

EXAMPLE 4.10 Frequency reuse distance, D

Determine the distance from the nearest cochannel cell for a cell having a radius of 0.64 km and a cochannel reuse factor of 12.

Solution

The radius of a cell, R = 0.64 km (given)

The cochannel reuse factor, q = 12 (given)

To determine the distance from the nearest cochannel cell, D

We know that q = D R,

Or, D = q × R

Therefore, D = 12 × 0.64 km = 7.68 km

Hence, the distance from the nearest cochannel cell D = 7.68 km

Thus, the important parameters of the network designed on cellular approach are of

• Reuse pattern, K
• Reuse distance, D
• Frequency reuse factor, q

- The concept of frequency reuse when applied, permits the system to meet the important objective of serving a large area, while using a relatively small frequency spectrum. But if the network is not designed properly, serious interferences may occur.
- To minimise interference, there must be adequate spatial separation between cells that use the same frequencies and the cells that use adjacent channel frequencies.
- The frequency assignment depends on the channel bandwidth, modulation scheme adopted, reuse factor and the carrier-to-interference ratio requirements.

Problem 6

EXAMPLE 4.11 Frequency reuse ratio, q

Determine the frequency reuse ratio for a cell radius of 0.8 km separated from the nearest co channel cell by a distance of 6.4 km.

R 1

R 1

Fig. 4.14 Frequency reuse ratio q = D ⁄ R

The real power of the cellular concept is that interference is not related to the absolute distance between cells but to the ratio of the distance between cochannel (same frequency) cells to the cell radius.

Solution

The radius of a cell, R = 0.8 km (given)

The distance between nearest cochannel cells, D = 6.4 km (given)

To determine the frequency reuse ratio, q

We know that q = D R

Or, q = 6.4 0.8 = 8

Hence, the frequency reuse ratio for given parameters q = 8

- The frequency reuse method is useful for increasing the efficiency of spectrum usage but results in cochannel interference because the same frequency channel is used repeatedly in different cochannel cells in a service area.
- In this situation, the received signal quality is affected by the amount of radio coverage area as well as the cochannel interference.
- The cochannel interference is caused due to the reuse of the same carrier frequency at different geographical locations.
- Because cochannel interfering signals are amplified, processed and detected in the same manner as the desired signal, the receiver is particularly vulnerable to these emissions.
- Thus, cochannel interference may either desensitise the receiver or override or mask the desired signal.

- The cochannel interference can then be measured by selecting any one channel (as one channel represents all the channels) and transmitting on that channel at all cochannel sites.
- In a fully equipped hexagonal-shaped cellular system, there are always six cochannel interfering cells in the first tier. Figure 4.17 depicts a typical field measurement test set-up 1 to measure cochannel interference at the mobile unit, in which the mobile unit is moving in its serving cell.
- Let the symbol C, I, and N denote respectively the power of the desired signal, the power of the cochannel interference, and the power of the noise at the output of the receiver demodulator. Cochannel interference can be experienced both at the cell-site and at mobile units in the serving cell. If the interference is much greater then the carrier to interference ratio C/I at the mobile units caused by the six interfering cell-sites is (on the average) the same as the C/I received at the serving cell site caused by interfering mobile units in the six cells.
- According to the reciprocity theorem and the statistical summation of radio propagation, the two C/I values can be very close.

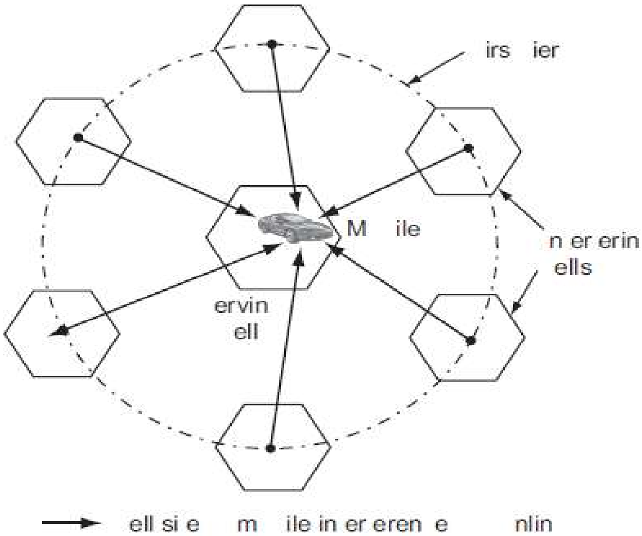


Fig. 4.17 *Test 1: Cochannel interference measurement at the mobile unit*

A channel-scanning mobile receiver records three received signals while moving in any one cochannel cell, under the following conditions:

- When only the serving cell transmits (signal recorded is termed as C)
- Cell-sites of all six cochannel cells only transmit (signal recorded is termed as I)
- No transmission by any cell-site (signal recorded is termed as N)

MODULE 5

GLOBAL SYSTEM FOR MOBILE

GSM Network Architecture, GSM Signalling protocol architecture, Identifiers used in GSM system, GSM Channels, Frame structure for GSM, GSM Call procedures, GSM hand-off Procedures, GSM Services and features

Global System for Mobile Communications (GSM) services are a standard collection of applications and features available to mobile phone subscribers all over the world. The GSM standards are defined by the 3GPP collaboration and implemented in hardware and software by equipment manufacturers and mobile phone operators. The common standard makes it possible to use the same phones with different companies' services, or even roam into different countries. GSM is the world's most dominant mobile phone standard.

The design of the service is moderately complex because it must be able to locate a moving phone anywhere in the world, and accommodate the relatively small battery capacity, limited input/output capabilities, and weak radio transmitters on mobile devices.

In order to gain access to GSM services, a user needs three things:

- A billing relationship with a mobile phone operator. This is usually either where services are paid for in advance of them being consumed (prepaid), or where bills are issued and settled after the service has been consumed (postpaid).
- A mobile phone that is GSM compliant and operates at the same frequency as the operator. Most phone companies sell phones from third-party manufacturers.
- A Subscriber Identity Module (SIM) card, which is activated by the operator once the billing relationship, is established. After activation the card is then programmed with the subscriber's Mobile Subscriber Integrated Services Digital Network Number (MSISDN) (the telephone number). Personal information such as contact numbers of friends and family can also be stored on the SIM by the subscriber.

After subscribers sign up, information about their identity (telephone number) and what services they are allowed to access are stored in a "SIM record" in the Home Location Register (HLR).

Once the SIM card is loaded into the phone and the phone is powered on, it will search for the nearest mobile phone mast (also called a Base Transceiver Station/BTS) with the strongest signal in the operator's frequency band. If a mast can be successfully contacted, then there is said to be coverage in the area. The phone then identifies itself to the network through the control channel. Once this is successfully completed, the phone is said to be attached to the network.

The key feature of a mobile phone is the ability to receive and make calls in any area where coverage is available. This is generally called roaming from a customer perspective, but also called visiting when describing the underlying technical process. Each geographic area has a database called the Visitor Location Register (VLR), which contains details of all the mobiles currently in that area. Whenever a phone attaches, or visits, a new area, the Visitor Location Register must contact the Home Location Register to obtain the details for that phone. The Current cellular location of the phone (i.e., which BTS it is at) is entered into the VLR record and will be used during a process called paging when the GSM network wishes to locate the mobile phone.

GSM Band	Uplink Frequency	Downlink Frequency
P-GSM 900 ARFCN=1...124	890 - 915 MHz (ARFCN-1) × 0.2 MHz + 890.2 MHz	935 - 960 MHz Uplink frequency + 45 MHz
E-GSM 900 ARFCN=975...1023	880 - 890 MHz (ARFCN = 0 = 890 MHz) (ARFCN-975) × 0.2 MHz + 890 MHz	925 - 935 MHz Uplink frequency + 45 MHz
R-GSM 900 ARFCN=955...1023	876 - 890 MHz (ARFCN-1023) × 0.2 MHz + 890 MHz	921 - 935 MHz Uplink frequency + 45 MHz
GSM 1800 ARFCN=512...885	1710 - 1785 MHz (ARFCN-512) × 0.2 MHz + 1710.2 MHz	1805 - 1880 MHz Uplink frequency + 95 MHz
GSM 1900 ARFCN=512...810	1850 - 1910 MHz (ARFCN-512) × 0.2 MHz + 1850.2 MHz	1930 - 1990 MHz Uplink frequency + 90 MHz
GSM 450 ARFCN=259...293	450.4 - 457.6 MHz (ARFCN-259) × 0.2 MHz + 450.6 MHz	460.4 - 467.6 MHz Uplink frequency + 10 MHz
GSM 480 ARFCN=306...340	478.8 - 486 MHz (ARFCN-306) × 0.2 MHz + 478.8 MHz	488.8 - 496 MHz Uplink frequency + 10 MHz

Every SIM card contains a secret key, called the Ki, which is used to provide authentication and encryption services. This is useful to prevent theft of service, and also to prevent "over the air" snooping of a user's activity. The network does this by utilizing the Authentication Center and is accomplished without transmitting the key directly.

Every GSM phone contains a unique identifier (different from the phone number), called the International Mobile Equipment Identity (IMEI). This can be found by dialing *#06#. When a phone contacts the network, its IMEI may be checked against the Equipment Identity Register to locate stolen phones and facilitate monitoring.

Time Division Multiple Access (TDMA):

It can be easily adapted to the transmission of data and voice communication. TDMA offers the ability to carry data rates of 64 kbps to 120 Mbps (expandable in multiples of 64 kbps). This enables operators to offer personal communication-like services including fax, voice band data, and short message services (SMSs) as well as bandwidth-intensive applications such as multimedia and videoconferencing. It will not experience interference from other simultaneous transmissions Unlike spread-spectrum techniques which can suffer from interference among the users all of whom are on the same frequency band and transmitting at the same time, TDMA's technology, which separates users in time, ensures that they will not TDMA is the only technology that offers an efficient utilization of hierarchical cell structures (HCSs) offering pico, micro, and macro cells. HCSs allow coverage for the system to be tailored to support specific traffic and service needs. By using this approach, system capacities of more than 40-times AMPS can be achieved in a cost-efficient way. TDMA allows service compatibility with the use of dual-mode handsets because of its inherent compatibility with FDMA analog systems.

2.2 GSM Network and System Architecture

Mobile Station (MS)

Mobile Equipment (ME)

Subscriber Identity Module (SIM)

Base Station Subsystem (BSS)

Base Transceiver Station (BTS)

Base Station Controller (BSC)

Network Switching Subsystem (NSS)

Mobile Switching Center (MSC)

Home Location Register (HLR)

Visitor Location Register (VLR)

Authentication Center (AUC)

Equipment Identity Register (EIR)

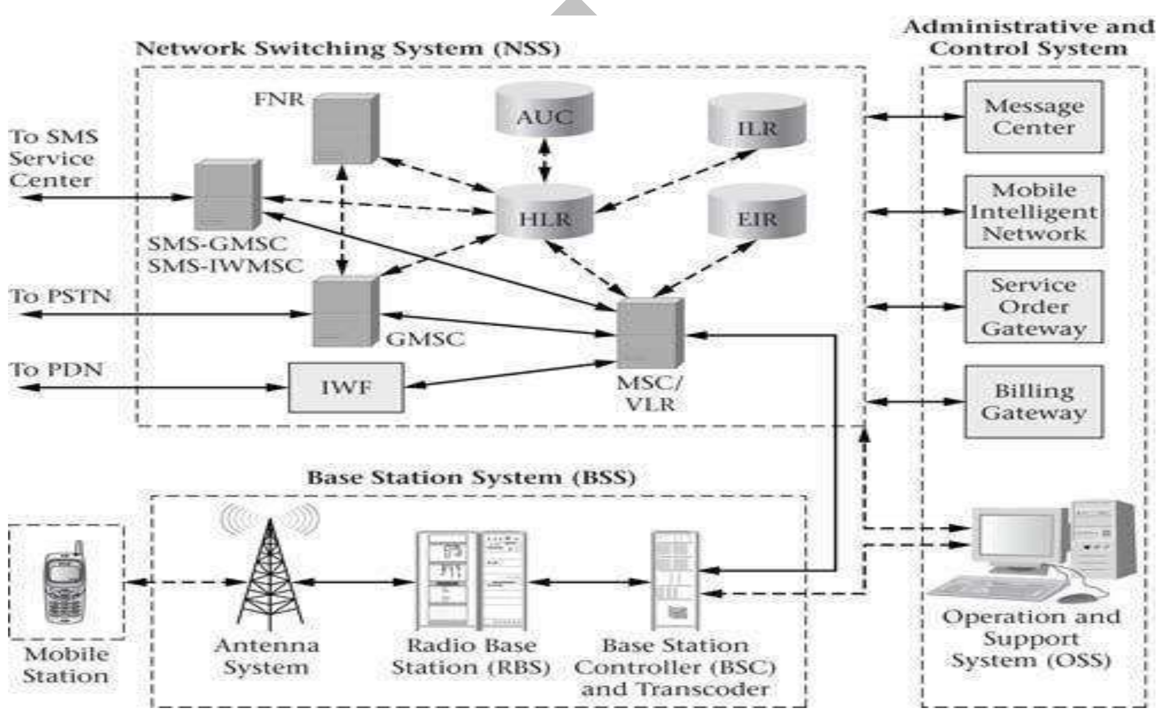


Fig 4.1 components of GSM network

a. Mobile Station (MS)

The Mobile Station is made up of two entities:

Mobile Equipment (ME):

- Portable, vehicle mounted, hand held device

- Uniquely identified by an IMEI (International Mobile Equipment Identity)
- Voice and data transmission
- Monitoring power and signal quality of surrounding cells for optimum handover
- Power level : 0.8W – 20 W
- 160 character long SMS.

Subscriber Identity Module (SIM)

- Smart card contains the International Mobile Subscriber Identity (IMSI)
- Allows user to send and receive calls and receive other subscribed services
- Protected by a password or PIN
- Can be moved from phone to phone – contains key information to activate the phone

b. **Base Station Subsystem (BSS):** Base Station Subsystem is composed of two parts that communicate across the standardized Abis interface allowing operation between components made by different suppliers.

Base Transceiver Station (BTS): It is responsible for following functions

- Encodes, encrypts, multiplexes, modulates and feeds the RF signals to the antenna.
- Communicates with Mobile station and BSC
- Consists of Transceivers (TRX) units

Base Station Controller (BSC): It is responsible for following functions

- Manages Radio resources for BTS
- Assigns Frequency and time slots for all MS's in its area
- Handles call set up
- Handover for each MS
- It communicates with MSC and BTS

c. **Network Switching Subsystem(NSS)**

Mobile Switching Center (MSC) :

- Heart of the network
- Manages communication between GSM and other networks
- Billing information and collection
- Mobility management
- Registration
- Location Updating
- Inter BSS and inter MSC call handoff

Home Location Register (HLR):

- Stores information about each subscriber that belongs to it MSC in permanent and temporary fashion.
- As soon as mobile subscriber leaves its current local area, the information in the HLR is updated.
- Database contains IMSI, MSISDN, prepaid/ postpaid, roaming restrictions, supplementary services.

Visitor Location Register (VLR):

- Temporary database which updates whenever new MS enters its area, by HLR database.
- Assigns a TMSI to each MS entering the VLR area which keeps on changing.
- Controls those mobiles roaming in its area.
- Database contains IMSI, MSISDN, Location Area, authentication key.

Authentication Centre (AUC):

- Contains the algorithms for authentication as well as the keys for encryption.
- Protects network operators from fraud.
- Situated in special protected part of the HLR.

Equipment Identity Register (EIR):

- Stores all devices identifications registered for this network.
- Database that is used to track handsets using the IMEI(International Mobile Equipment Identity)
- Prevents calls from stolen, unauthorized or defective mobile devices

Operation and Maintenance Centre (OMC):

- The centralized operation of the various units in the system and functions needed to maintain the subsystems.
- Dynamic monitoring and controlling of the network.
- Functions :
    - configuration management
    - fault report and alarm handling
    - performance supervision/management
    - storage of system software and data

2.3 GSM network interfaces and protocols:

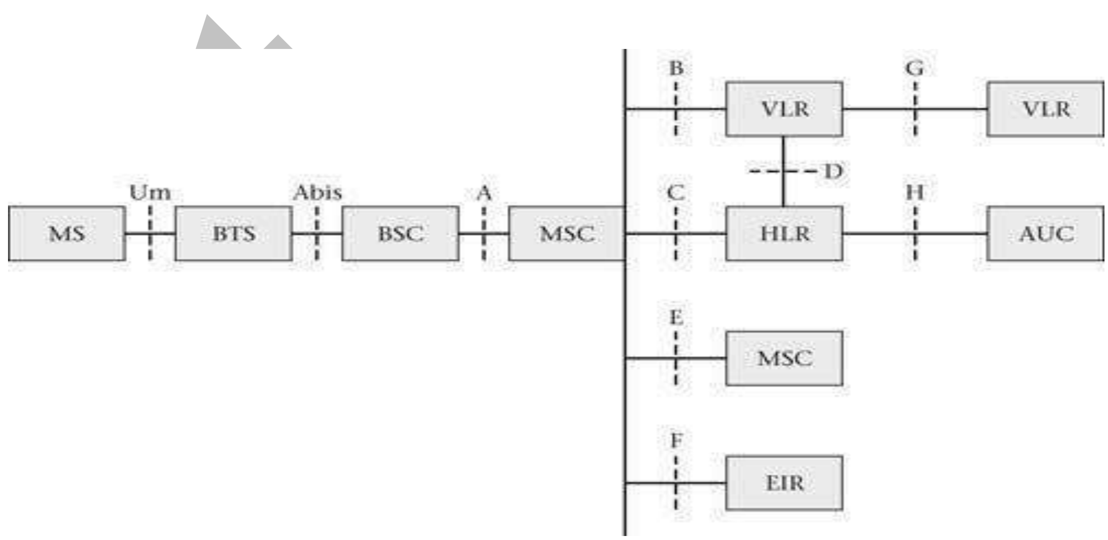


Fig 4.2 interfaces in GSM

GSM network interfaces and protocols:

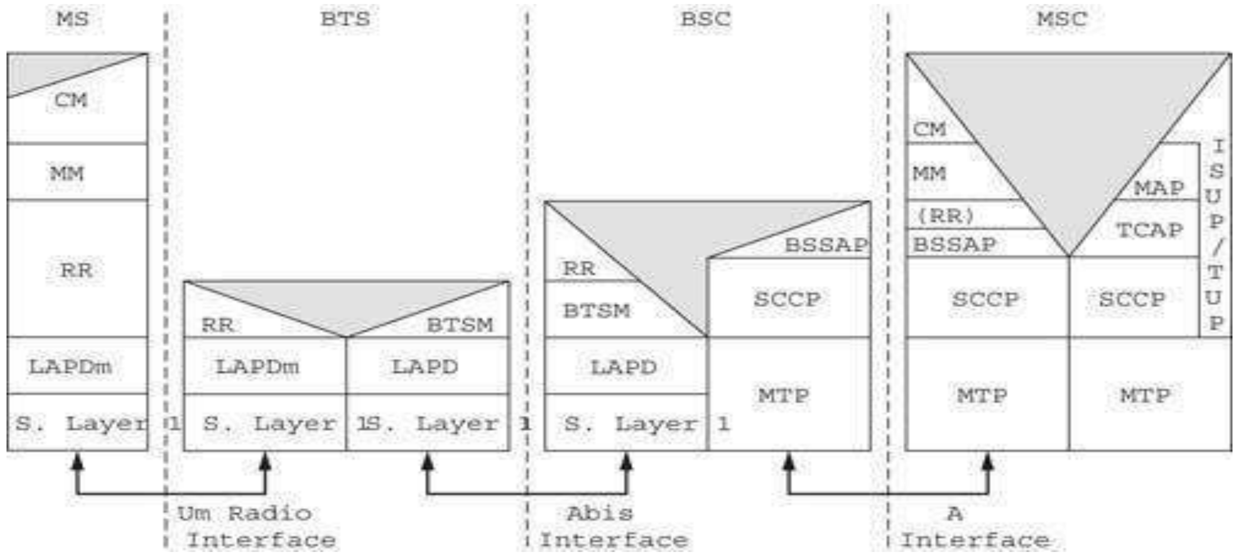


Fig 4.3 GSM network interfaces and protocols

The network structure is defined within the GSM standards. Additionally each interface between the different elements of the GSM network is also defined. This facilitates the information interchanges can take place. It also enables to a large degree that network elements from different manufacturers can be used. However as many of these interfaces were not fully defined until after many networks had been deployed, the level of standardization may not be quite as high as many people might like

1. **Um interface:** The "air" or radio interface standard that is used for exchanges between a mobile (ME) and a base station (BTS / BSC). For signaling, a modified version of the ISDN LAPD, known as LAPDm is used.
2. **Abis interface:** This is a BSS internal interface linking the BSC and a BTS, and it has not been totally standardized. The Abis interface allows control of the radio equipment and radio frequency allocation in the BTS.
3. **A interface** The A interface is used to provide communication between the BSS and the MSC. The interface carries information to enable the channels, timeslots and the like to be allocated to the mobile equipment being serviced by the BSSs. The messaging required within the network to enable handover etc to be undertaken is carried over the interface.
4. **B interface** The B interface exists between the MSC and the VLR . It uses a protocol known as the MAP/B protocol. As most VLRs are collocated with an MSC, this makes the interface purely an "internal" interface. The interface is used whenever the MSC needs access to data regarding a MS located in its area.
5. **C interface** The C interface is located between the HLR and a GMSC or a SMS-G. When a call originates from outside the network, i.e. from the PSTN or another mobile network it as to pass through the gateway so that routing information required to complete the call may be gained. The protocol used for communication is MAP/C, the letter "C" indicating that the protocol is used for the "C" interface. In addition to this, the MSC may optionally forward billing information to the HLR after the call is completed and cleared down.

6. **D interface** The D interface is situated between the VLR and HLR. It uses the MAP/D protocol to exchange the data related to the location of the ME and to the management of the subscriber.
7. **E interface:** The E interface provides communication between two MSCs. The E interface exchanges data related to handover between the anchor and relay MSCs using the MAP/E protocol.
8. **F interface:** The F interface is used between an MSC and EIR. It uses the MAP/F protocol. The communications along this interface are used to confirm the status of the IMEI of the ME gaining access to the network.
9. **G interface:** The G interface interconnects two VLRs of different MSCs and uses the MAP/G protocol to transfer subscriber information, during e.g. a location update procedure.
10. **H interface:** The H interface exists between the MSC the SMS-G. It transfers short messages and uses the MAP/H protocol.
11. **I interface:** The I interface can be found between the MSC and the ME. Messages exchanged over the I interface are relayed transparently through the BSS.
12. Although the interfaces for the GSM cellular system may not be as rigorously defined as many might like, they do at least provide a large element of the definition required, enabling the functionality of GSM network entities to be defined sufficiently.

2.3 GSM Channel Concept:

- Time division multiple access
- Frames Multi frames

A single GSM RF carrier can support up to eight MS subscribers simultaneously. Each channel occupies the carrier for one eighth of the time.
This is a technique called Time Division Multiple Access. Time is divided into discrete periods called Time slots. The timeslots are arranged in sequence and are conventionally numbered 0 to 7. Each repetition of this sequence is called a TDMA frames. Each MS telephone call occupies one timeslot (0â€“7) within the frame until the call is terminated, or a handover occurs.

The TDMA frames are then built into further frame structures according to the type of channel. We shall later examine how the information carried by the air interface builds into frames and multi-frames and discuss the associated timing. For such a system to work correctly, the timing of the transmissions to and from the mobiles is critical. The MS or Base Station must transmit the information related to one call at exactly the right moment, or the timeslot will be missed. The information carried in one timeslot is called a access burst . Each data burst, occupying its allocated timeslot within successive TDMA frames, provides a single GSM physical channel carrying a varying number of logical channels between the MS and BTS.

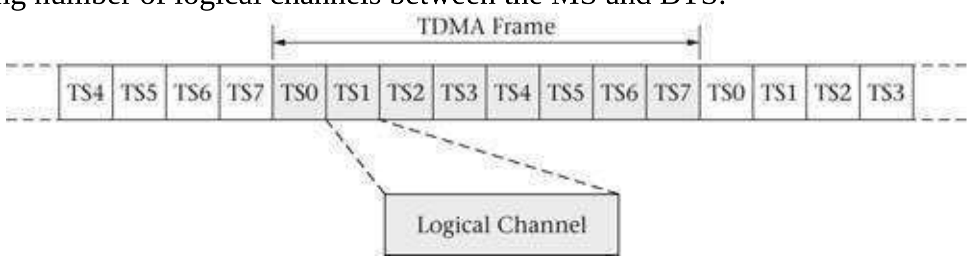


Fig 4.4 TDMA time frame structure

GSM Channel Concept:

Logical channels:

- Carry either sub scriber traffic or signaling and control information to facilitate subscriber mobility.
- Presently, there are three types of traffic channels (TCHS).
- The full-rate traffic channel (TCH/F or Bm) carries one conversation by using one timeslot. The transmitted voice signal is encoded at a 13-kbps rate, but it is sent with additional overhead bits. This information plus additional channel overhead bits yields a final channel data rate of 22.8 kbps. The full-rate traffic channel may also carry data at rates of 14.4, 9.6, 4.8, and 2.4 kbps.
- The half-rate traffic channel (TCH/H or Lm) carries voice encoded at 6.5 kbps or data at rates of 4.8 or 2.4 kbps. With additional overhead bits, the total data rate for TCH/H becomes 11.4 kbps. Therefore, two conversations or a conversation and a data transfer or two data transfers may be transmitted over one channel at the same time.

- Enhanced full-rate (EFR) traffic encodes voice at a 12.2-kbps rate and like TCH/F adds overhead bits to yield a 22.8 kbps channel data rate. The EFR channel may also transmit data at the TCH/F rates. More will be said about these channels later. The signaling and control channels consist of three channel sub categories: broadcast channels, common control channels, and dedicated control channels. The function of these channels will be explained in more detail next. Broadcast channels

Broadcast control channel

- The GSM cellular system uses broadcast channels (BCHS) provide information to the mobile station about various system parameters and also information about the location area identity (LAI).
- The three types BCHS are
- Broadcast control channel: It contains information that needed by MS concerning the cell that it is attached to in order for the MS to be able to start making or receiving calls, or to start roaming.
- Frequency correction channel: It transmits bursts of zeros (this is an unmodulated carrier signal) to the MS. This signaling is done for two reasons: the MS can use this signal to synchronize itself to the correct frequency and the MS can verify that this is the BCCH carrier.
- Synchronization channel: It transmits the required information for the MS to synchronize itself with the timing within a particular cell. By listening to the SCH, the MS can learn about the frame number in this cell and about the BSIC of the BTS it is attached to.
- Using the information transmitted over these three BCHS, the MS can tune to particular base transceiver system (BTS) and synchronize its timing with the frame structure and timing that cell.
- Each time the MS attaches to new BTS must listen these three BCHS.

Common control channels:

- The common control channels (CCCHS) provide paging messages the MS and a means which the mobile can request signaling channel that it can use to contact the network.

- The three CCCHs are
- Paging channel: It is used by the system to send paging messages to mobiles attached to cell. The mobile will paged whenever the network has an incoming call ready for mobile or some type of message (e.g., short message multimedia message) to deliver to the mobile. The information transmitted the PCH will consist of paging message and the mobile's identity number.
- Random access channel: It is used by the mobile to respond a paging message. If the mobile receives page on the PCH, it will reply on the RACH with request for signaling channel.
- Access Grant channel: It is used by the network to assign a signaling channel to the MS. After the mobile requests a signaling channel over the RACH, the network will assign a channel to the mobile by transmitting this information over the AGCH. The AGCH is only transmitted in the downlink direction.

Dedicated control channels:

- These dedicated channels are used for specific call setup, handover, measurement, and short message delivery functions.
- The four DCCHs are
- Standalone dedicated control channel : Both the mobile station and the BTS switch over to the network-assigned stand-alone dedicated control channel (SDCCH) that is assigned over the access grant channel in response to the mobile's request that has been transmitted over the random access channel.
- Slow associated control channel: It is used to transmit information about measurements made by the MS or instructions from the BTS about the mobile's parameters of operation. In the uplink direction the mobile sends measurements of the received signal strength from its own BTS and those of neighboring BTSS. In the downlink direction, the MS receives information from the BTS about the mobile's output power level and the timing advance that the
- Fast associated control channel: It is used to facilitate the handover operation in a GSM system. If handover is required, the necessary handover signaling information is transmitted instead of a 20-ms segment of speech over the TCH. This operation is known as "stealing mode" since the time allotted for the voice conversation is stolen from the system for a short period. The sub scriber is usually not aware of this loss of speech since the speech coder in the mobile simply repeats the last received voice block during this process.
- Cell Broadcast channel: It is used to deliver short message service in the downlink direction. It uses the same physical channel as the SDCCH.

Timeslots and TDMA frames

TDMA frames

TDMA multiframes: Hyperframes, super frame, Multi frames (26+51 Frames)
Timeslot bursts : Normal Burst, Frequency Correction burst, Synchronization burst, Access burst, Dummy burst.

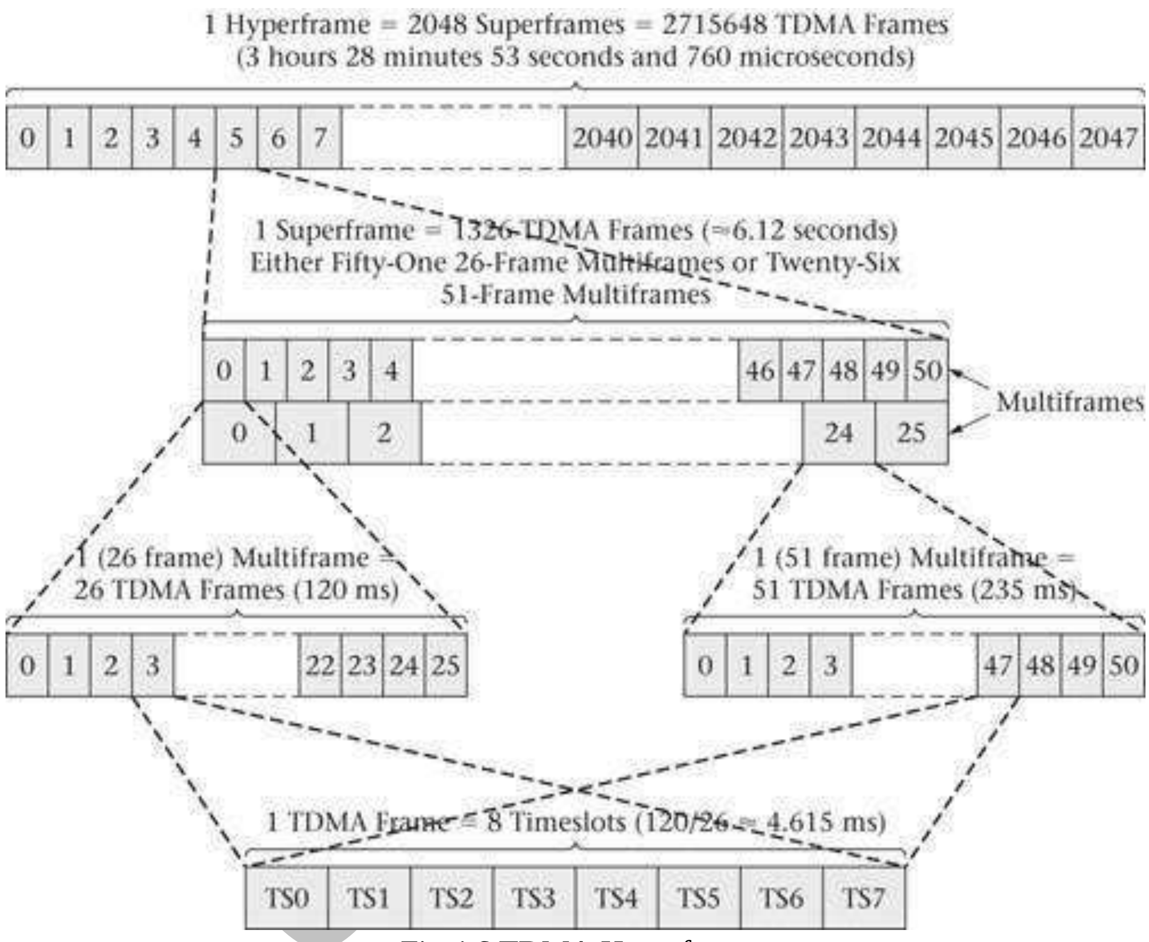


Fig 4.6 TDMA Hyperframe structure

A hyperframe is a multiframe sequence that is composed of 2048 superframes and is largest time interval in the GSM system (3 hours, 28 minutes, 53 seconds). Every time slot during a hyperframe has a sequential number (represented by an 11 bit counter) that is composed of a frame number and a time slot number. This counter allows the hyperframe to synchronize frequency hopping sequence, encryption processes for voice privacy of subscribers' conversations. The hyperframe in an IS-136 TDMA system consists of 192 frames.

The basic GSM frame defines the structure upon which all the timing and structure of the GSM messaging and signalling is based. The fundamental unit of time is called a burst period and it lasts for approximately 0.577 ms (15/26 ms). Eight of these burst periods are grouped into what is known as a TDMA frame. This lasts for approximately 4.615 ms (i.e.120/26 ms) and it form the basic unit for the definition of logical channels. One physical channel is one burst period allocated in each TDMA frame.

In simplified terms the base station transmits two types of channels, namely traffic and control. Accordingly, the channel structure is organised into two different types of frame, one for the traffic on the main traffic carrier frequency, and the other for the control on the beacon frequency.

GSM Multiframe:

The GSM frames are grouped together to form multiframes and in this way it is possible to establish a time schedule for their operation and the network can be synchronised.

There are several GSM multiframe structures:

- **Traffic multiframe**: The Traffic Channel frames are organized into multi frames consisting of 26 bursts and taking 120 ms. In a traffic multi frame, 24 bursts are used for traffic. These are numbered 0 to 11 and 13 to 24. One of the remaining bursts is then used to accommodate the SACCH, the remaining frame remaining free. The actual position used alternates between position 12 and 25.
- **Control multiframe:** the Control Channel multi frame that comprises 51 bursts and occupies 235.4 ms. This always occurs on the beacon frequency in time slot zero and it may also occur within slots 2, 4 and 6 of the beacon frequency as well. This multiframe is subdivided into logical channels which are time-scheduled.

GSM Superframe

Multi frames are then constructed into super frames taking 6.12 seconds. These consist of 51 traffic multi frames or 26 control multi frames. As the traffic multi frames are 26 bursts long and the control multi frames are 51 bursts long, the different number of traffic and control multiframes within the super frame, brings them back into line again taking exactly the same interval.

GSM Hyperframe

Above this 2048 super frames (i.e. 2 to the power 11) are grouped to form one hyper frame which repeats every 3 hours 28 minutes 53.76 seconds. It is the largest time interval within the GSM frame structure.Within the GSM hyper frame there is a counter and every time slot has a unique sequential number comprising the frame number and time slot number. This is used to maintain synchronization of the different scheduled operations with the GSM frame structure. These include functions such as:

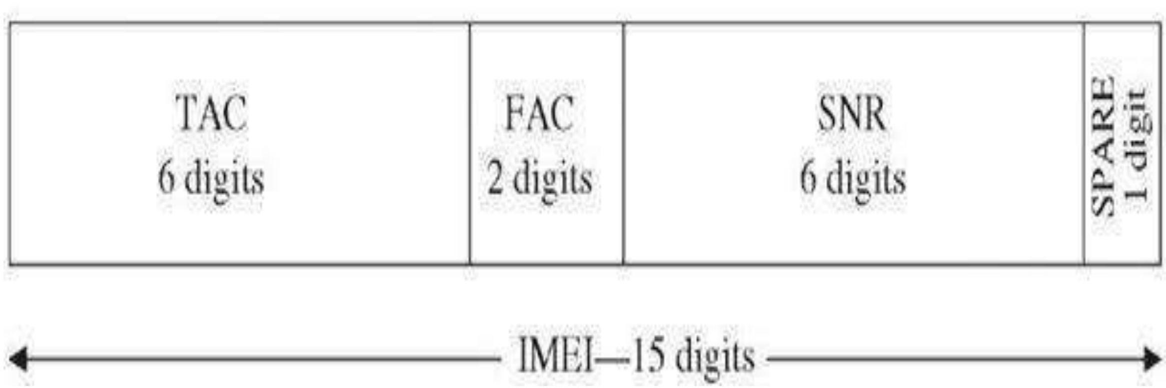
- **Frequency hopping**: Frequency hopping is a feature that is optional within the GSM system. It can help reduce interference and fading issues, but for it to work, the transmitter and receiver must be synchronised so they hop to the same frequencies atthe same time.
- **Encryption**: The encryption process is synchronized over the GSM hyper frame period where a counter is used and the encryption process will repeat with each hyper frame. However, it is unlikely that the cellphone conversation will be over 3 hours and accordingly it is unlikely that security will be compromised as a result.

GSM SYSTEM OPERATIONS

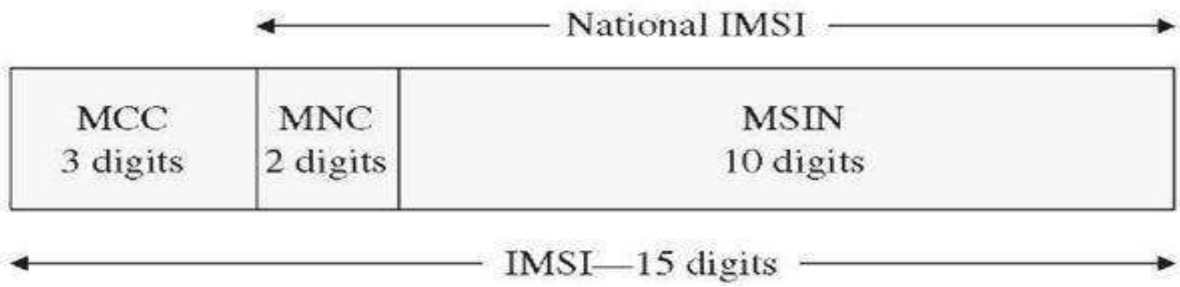
GSM IDENTIFIERS:
GSM treats the users and the equipment in different ways.
Phone numbers, subscribers, and equipment identifiers are some of the known ones.
There are many other identifiers that have been well-defined, which are required for the subscriber's mobility management and for addressing the remaining network elements.

a. **International Mobile Station Equipment GSM Identity (IMEI)**

The International Mobile Station Equipment Identity (IMEI) looks more like a serial number which distinctively identifies a mobile station internationally. This is allocated by the equipment manufacturer and registered by the network operator, who stores it in the Equipment Identity Register (EIR). By means of IMEI, one recognizes obsolete, stolen, or non-functional equipment.
Following are the parts of IMEI −
Type Approval Code (TAC) − 6 decimal places, centrally assigned.
Final Assembly Code (FAC) − 6 decimal places, assigned by the manufacturer.
Serial Number (SNR) − 6 decimal places, assigned by the manufacturer.
Spare (SP) − 1 decimal place.
Thus, **IMEI = TAC + FAC + SNR + SP.**
It uniquely characterizes a mobile station and gives clues about the manufacturer and the date of manufacturing.



b. **International Mobile Subscriber Identity (IMSI)**
Every registered user has an original International Mobile Subscriber Identity (IMSI) with a valid IMEI stored in their Subscriber Identity Module (SIM).
IMSI comprises of the following parts−
Mobile Country Code (MCC) − 3 decimal places, internationally standardized.
Mobile Network Code (MNC) − 2 decimal places, for unique identification of mobile network within the country.
Mobile Subscriber Identification Number (MSIN) − Maximum 10 decimal places, identification number of the subscriber in the home mobile network.



c. **Mobile Subscriber ISDN Number (MSISDN)**

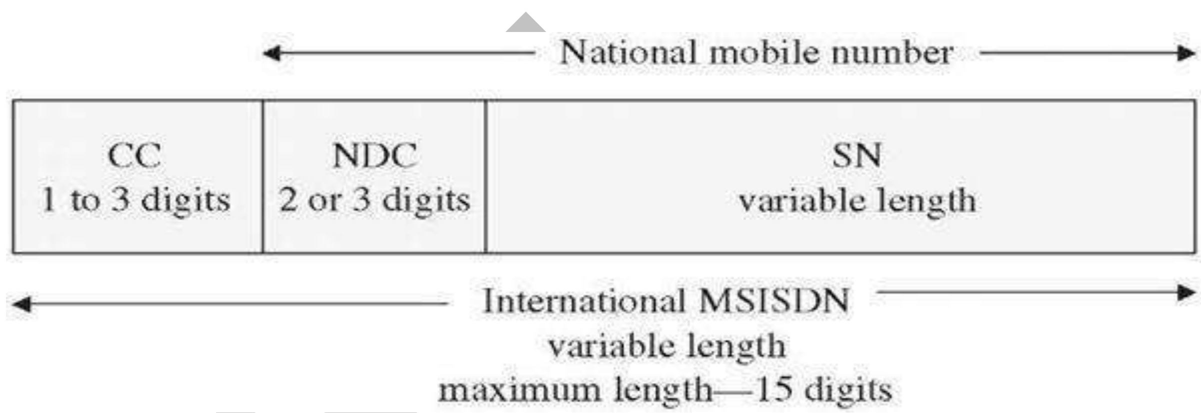
The authentic telephone number of a mobile station is the Mobile Subscriber ISDN Number (MSISDN). Based on the SIM, a mobile station can have many MSISDNs, as each subscriber is assigned with a separate MSISDN to their SIM respectively.

Listed below is the structure followed by MSISDN categories, as they are defined based on international ISDN number plan

Country Code (CC) − Up to 3 decimal places.

National Destination Code (NDC) − Typically 2-3 decimal places.

Subscriber Number (SN) − Maximum 10 decimal places.



d. **Mobile Station Roaming Number (MSRN)**

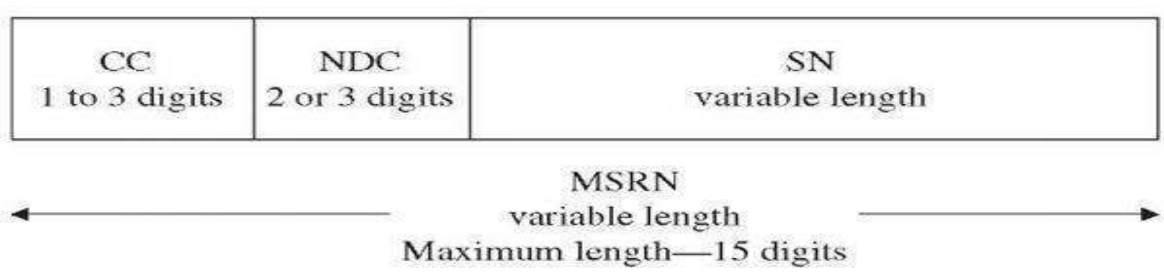
Mobile Station Roaming Number (MSRN) is an interim location dependent ISDN number, assigned to a mobile station by a regionally responsible Visitor Location Register (VLA). Using MSRN, the incoming calls are channeled to the MS.

The MSRN has the same structure as the MSISDN.

Country Code (CC) − of the visited network.

National Destination Code (NDC) − of the visited network.

Subscriber Number (SN) − in the current mobile network.



e. Location Area Identity (LAI)

Within a PLMN, a Location Area identifies its own authentic Location Area Identity (LAI). The LAI hierarchy is based on international standard and structured in a unique format as mentioned below

Country Code (CC) − 3 decimal places.

Mobile Network Code (MNC) − 2 decimal places.

Location Area Code (LAC) − maximum 5 decimal places or maximum twice 8 bits coded in hexadecimal (LAC < FFFF).

f. Temporary Mobile Subscriber Identity (TMSI)

Temporary Mobile Subscriber Identity (TMSI) can be assigned by the VLR, which is responsible for the current location of a subscriber. The TMSI needs to have only local significance in the area handled by the VLR. This is stored on the network side only in the VLR and is not passed to the Home Location Register (HLR).

Together with the current location area, the TMSI identifies a subscriber uniquely. It can contain up to 4×8 bits.

g. Local Mobile Subscriber Identity (LMSI)

Each mobile station can be assigned with a Local Mobile Subscriber Identity (LMSI), which is an original key, by the VLR. This key can be used as the auxiliary searching key for each mobile station within its region. It can also help accelerate the database access. An LMSI is assigned if the mobile station is registered with the VLR and sent to the HLR. LMSI comprises of four octets (4x8 bits).

h. Cell Identifier (CI)

Using a Cell Identifier (CI) (maximum 2×8) bits, the individual cells that are within an LA can be recognized. When the Global Cell Identity (LAI + CI) calls are combined, then it is uniquely defined.

CALL SETUP

- Call setup within a GSM system consists of quite a few necessary operations. For either a mobile-originating call or a mole-terminating call the following ten operations need to be performed.
- For a mobile-terminating call it is necessary to perform an initial additional

operation as shown:

  - Interrogation (only for a mobile-terminating call)
  - Radio resource connection establishment.
  - Service request
  - Authentication
  - Ciphering mode setting
  - I MEI number check
  - T MS I allocation
  - Call initiation
  - Assignment of a traffic channel
  - User alerting signaling
  - Call accepted signaling

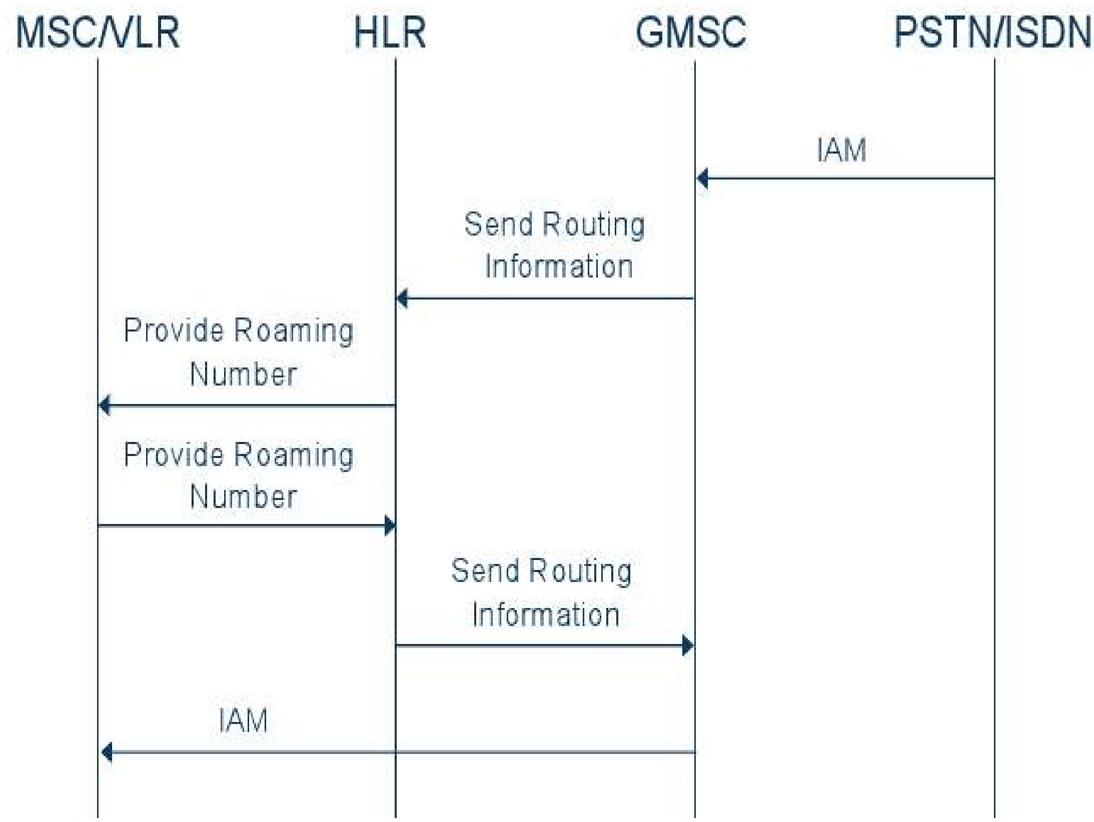
Interrogation Phase (MT):

- For the interrogation operation, one notes that the initial address message (IAM) comes outside the GSM network using ISUP/TUP protocols.

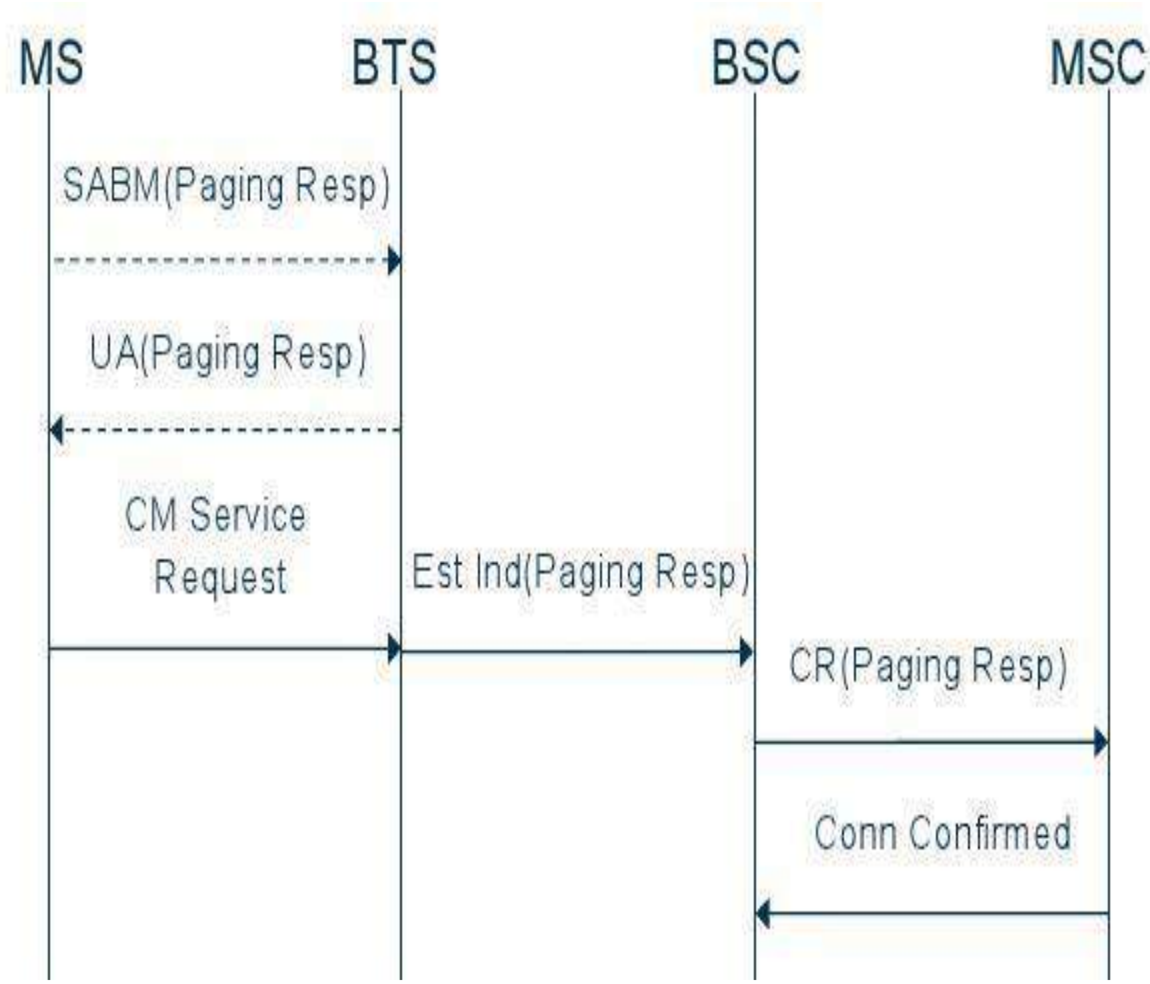
- In some vendors systems, the GMSC can send a request to the flexible numbering register (FNR) system node before being sent to the HLR .
- Also, for security operations, the subscriber data can

be simultaneously stored and updated in two HLRS.

- This built-in system redundancy assures successful in all but the most catastrophic disasters.
- In one final note about operation, one observes that in the last operation performed, the GSM system nodes (the MSC/VLR and the GMSC) use a non-MAP protocol to communicate with each other (i.e., the IAM message).



SERVICE REQUEST



The service request phase occurs as soon as the MS has tuned to the new channel assigned to it by the immediate assignment message sent during the radio resource connection phase.

 Figure shows these operations. At this time, a Layer 2 message known as set asynchronous balanced mode (SABM) is sent from the MS to the BTS. This Layer

2 message contains a Layer 3 message (i.e., the information field of the Layer 2 message contains the paging response message).

 Shortly thereafter, the BTS sends back to the MS a

GSM System Operations (Traffic Cases) : refer ppt from google classroom