Radio Wave Propagation

Block Diagram of Communication System

Fig.1 shows the block diagram of a general communication system, in which the different functional elements are represented by blocks.

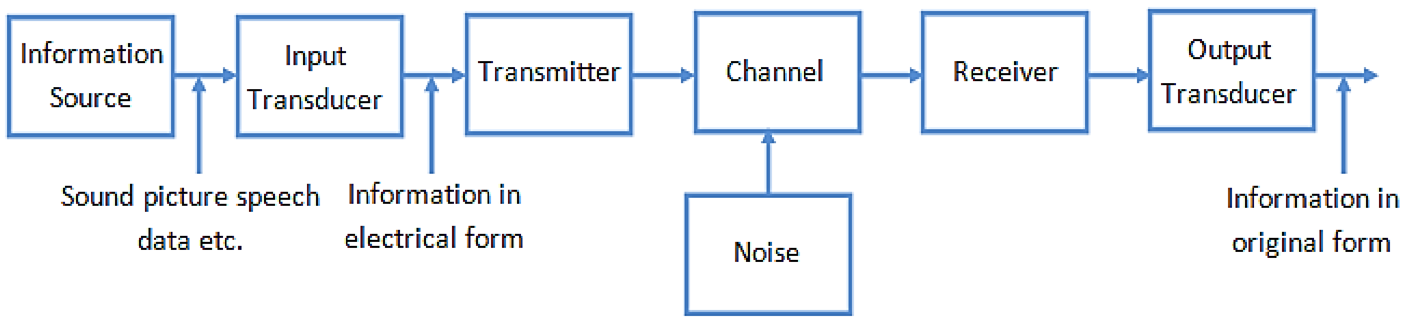
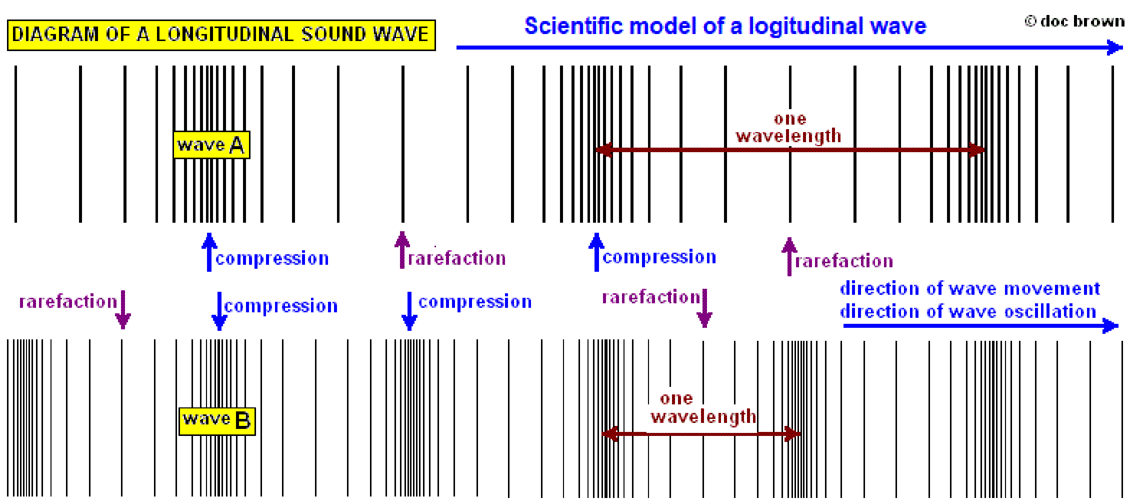
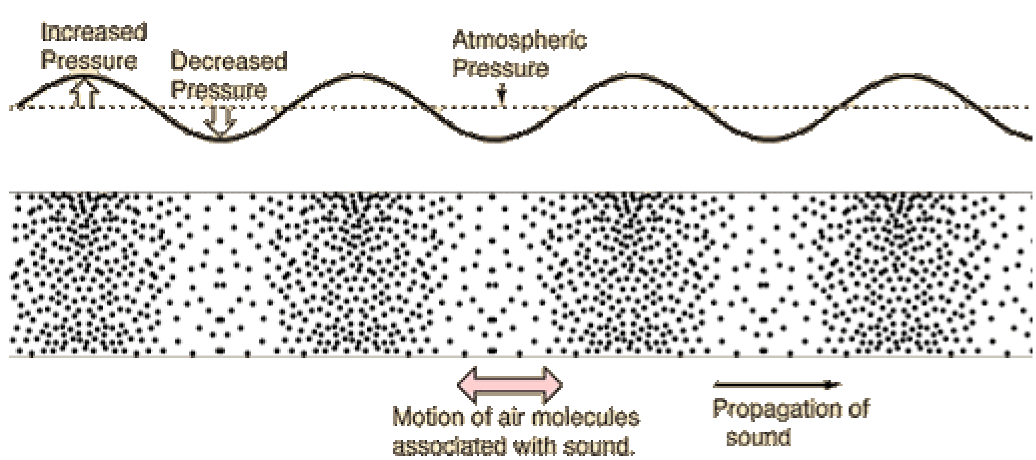


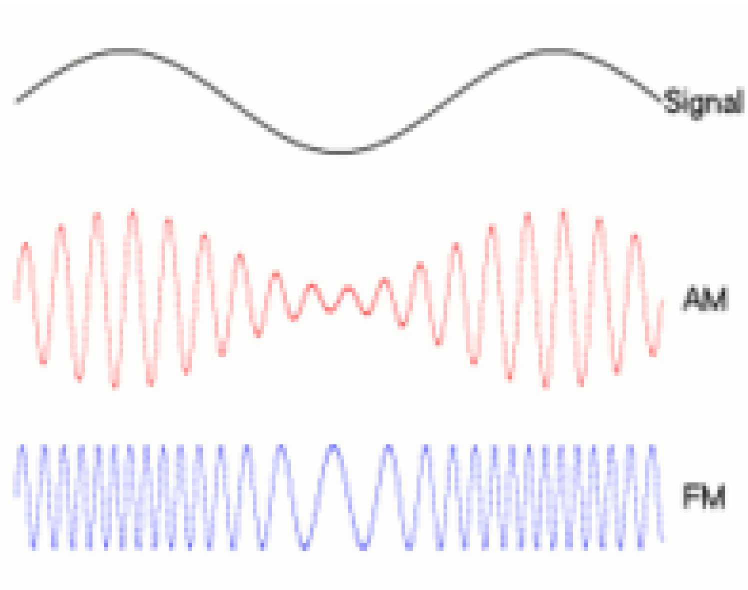
Fig 1





Md. Saifur Rahman, Lecturer, Department of Physics, University of Rajshahi / 3H /ECS/ Chapter-7 / 2020

Modulation



In electronics and telecommunications, modulation is the process of varying one or more properties of a periodic waveform, called the carrier signal, with a separate signal called the modulation signal that typically contains information to be transmitted. For example, the modulation signal might be an audio signal representing sound from a microphone, a video signal representing moving images from a video camera, or a digital signal representing a sequence of binary digits, a bitstream from a computer. The carrier is higher in frequency than the modulation signal. The purpose of modulation is to impress the information on the carrier wave, which is used to carry the information to another location. In radio communication the modulated carrier is transmitted through space as a radio wave to a radio receiver. Another purpose is to transmit multiple channels of information through a single communication medium, using frequency division multiplexing (FDM). For example, in cable television which uses FDM, many carrier signals carrying different television channels are transported through a single cable to customers. Since each carrier occupies a different frequency, the channels do not interfere with each other. At the destination end, the carrier signal is demodulated to extract the information bearing modulation signal.

In **analog modulation** an analog modulation signal is impressed on the carrier.
Examples are **amplitude modulation (AM)** in which the amplitude (strength) of the carrier wave is varied by the modulation signal, and **frequency modulation (FM)** in which the frequency of the carrier wave is varied by the modulation signal. These were the earliest types of modulation, and are used to transmit an audio signal representing sound, in AM and FM radio broadcasting.

- **AM**

  In amplitude modulation, the amplitude (signal strength) of the carrier wave is varied in proportion to that of the message signal, such as an audio signal.

- **FM**

  Frequency modulation (FM) is the encoding of information in a carrier wave by varying the instantaneous frequency of the wave.

- **PM**

  It encodes a message signal as variations in the instantaneous phase of a carrier wave. Phase modulation is one of the two principal forms of angle modulation, together with frequency modulation.

- **QAM**

Quadrature amplitude modulation (QAM) is the name of a family of digital modulation methods and a related family of analog modulation methods widely used in modern telecommunications to transmit information. It conveys two analog message signals, or two digital bit streams, by changing (modulating) the amplitudes of two carrier waves, using the amplitude-shift keying (ASK) digital modulation scheme or amplitude modulation (AM) analog modulation scheme. The two carrier waves of the same frequency are out of phase with each other by 90°, a condition known as orthogonality or quadrature. The transmitted signal is created by adding the two carrier waves together.

- **SM**

Space modulation is a radio amplitude modulation technique used in instrument landing systems (ILS) that incorporates the use of multiple antennas fed with various radio frequency powers and phases to create different depths of modulation within various volumes of three-dimensional airspace.

- **SSB**

Single-sideband modulation (SSB) or single-sideband suppressed-carrier modulation (SSB-SC) is a type of modulation used to transmit information, such as an audio signal, by radio waves. A refinement of amplitude modulation, it uses transmitter power and bandwidth more efficiently. Amplitude modulation produces an output signal the bandwidth of which is twice the maximum frequency of the original baseband signal. Single-sideband modulation avoids this bandwidth increase, and the power wasted on a carrier, at the cost of increased device complexity and more difficult tuning at the receiver.

More recent systems use **digital modulation**, which impresses a digital signal consisting of a sequence of binary digits (bits), a bitstream, on the carrier. In **frequency shift keying** (FSK) modulation, used in computer buses and telemetry, the carrier signal is periodically shifted between two frequencies that represent the two binary digits. In digital baseband modulation (line coding) used to transmit data in serial computer bus cables and wired LAN computer networks such as Ethernet, the voltage on the line is switched between two amplitudes (voltage levels) representing the two binary digits, 0 and 1, and the carrier (clock) frequency is combined with the data. A more complicated digital modulation method that employs multiple carriers, orthogonal frequency division multiplexing (OFDM), is used in WiFi networks, digital radio stations and digital cable television transmission.

The most fundamental digital modulation techniques are based on keying:

- **PSK** (phase-shift keying): a finite number of phases are used.
- **FSK** (frequency-shift keying): a finite number of frequencies are used.
- **ASK** (amplitude-shift keying): a finite number of amplitudes are used.
- **QAM** (quadrature amplitude modulation): a finite number of at least two phases and at least two amplitudes are used.

Necessity of Modulation

a) **Increase The Signal Strength.** The baseband signals transmitted by the sender are not capable of direct transmission. The strength of the message signal should be increased so that it can travel longer distances. This is where modulation is essential. The most vital need of modulation is to enhance the strength of the signal without affecting the parameters of the carrier signal.

b) **Wireless Communication System** Modulation has removed the necessity for using wires in the communication systems. It is because modulation is widely used in transmitting signals from one location to another with faster speed. Thus, the modulation technique has helped in enhancing wireless communication systems.

c) **Prevention of Message Signal from Mixing** Modulation and its types prevent the interference of the message signal from other signals. It is because a person sending a message signal through the phone cannot tell such signals apart. As a result, they will interfere with each other. However, by using carrier signals having a high frequency, the mixing of the signals can be prevented. Thus, modulation ensures that the signals received by the receiver are entirely perfect.

d) **Size of the Antenna:** Theory shows that in order to transmit a wave effectively the length of the transmitting antenna should be approximately equal to the wavelength of the wave or at least a quarter of the wavelength.

$$practical\ antenna\ size \approx \frac{wavelength}{4} = \frac{velocity}{4 \times frequency}$$

As the audio frequency range from $20\ Hz$ to $20\ KHz$ therefore if they are transmitted directly into space the length of the transmitting antenna would be extremely large. For instance, to radiate a frequency of $20\ KHz$ directly into space, we would need an antenna length of

$$= \frac{3 \times 10^8}{4 \times 20 \times 10^3} = 3750\ m$$

This is too long for the antenna to be constructed practically. On the other hand, if a carrier wave say $1000\ KHz$ is used to carry the signal, we need the antenna length of

$$= \frac{3 \times 10^8}{4 \times 1000 \times 10^3} = 75\ m$$

An antenna of this size can be easily constructed.

Types of Signals Used in The Modulation

- **Modulating Signal:** This is the signal that contains the message to be transmitted from the sender to the receiver and is called a message signal. Generally, the message signals are the band of low or high frequencies and are often called baseband signals. The message signals are the signals to be transmitted from the sender to the receiver. The frequency of the message signals to be sent is generally low. Thus, these signals undergo modulation to get correctly transmitted from one location to another.
- **Carrier Signal:** The other signal used in the process of modulation is the carrier signal that has high-frequency sinusoidal waves. The high-frequency carrier wave can travel much quicker as compared to the baseband signal. These signals have a specific frequency, amplitude, and phase, but no information. After modulation, carrier signals are used to transmit the signal to the receiver.
- **Modulated Signal:** After the modulation is done, the resultant signal refers to the modulated signal. This signal is the mixture of the carrier signal and message signal. (Image to be added soon) The diagram shows three types of signals, namely, message signal, carrier signal, and modulated signal that is the mixture of the message and carrier signal.

Amplitude Modulation

Let the carrier voltage be expressed as

$$e_c = E_c \cos(\omega_c t + \theta) \tag{1}$$

where E_c is the peak amplitude of carrier voltage, ω_c is the angular frequency of carrier wave and θ is the phase angle. Three cases may arise:

(i) When the amplitude E_c is varied in accordance with the modulating wave, the process is known as amplitude modulation; ω_c and θ remain constant.

(ii) When ω_c varies in accordance with the modulating wave, while E_c and θ remain constant, the process is known as frequency modulation.

(iii) When θ is varied in accordance with modulating wave, while E_c and ω_c remain constant, the process is known as phase modulation.

In amplitude modulation, the magnitude of the carrier wave is varied in accordance with the amplitude and frequency of the modulating voltage. Let the signal voltage be represented as

$$e_m = E_m \cos \omega_m t \tag{2}$$

The carrier frequency ω_c is much greater than the frequency of modulating signal or wave ω_m. The resulting modulated wave has the form

$$e = (E_c + K_a e_m) \cos(\omega_c t + \theta)$$

$$= (E_c + K_a E_m \cos \omega_m t) \cos(\omega_c t + \theta)$$

The amplitude factor $(E_c + K_a E_m \cos \omega_m t)$ represents the sinusoidal variation of the amplitude of the wave. K_a is the proportionality factor that determines the maximum variation in amplitude for a given signal voltage E_m. Phase θ is taken zero because it plays no part in this process. The above expression can be written as

$$e = (E_c + K_a E_m \cos \omega_m t) \cos \omega_c t \tag{3}$$

$$= E_c \left(1 + \frac{K_a E_m}{E_c} \cos \omega_m t\right) \cos \omega_c t$$

Where $m_a = \frac{K_a E_m}{E_c}$ is called modulation index. The modulation index when multiplied by 100 gives a percentage of modulation.

$$e = E_c (1 + m_a \cos \omega_m t) \cos \omega_c t$$

$$= E_c \cos \omega_c t + m_a E_c \cos \omega_m t \cos \omega_c t \qquad \because\ 2 \cos A \cos B = \cos(A+B) + \cos(A-B)$$

$$= E_c \cos \omega_c t + \frac{m_a E_c}{2} \cos(\omega_c + \omega_m) t + \frac{m_a E_c}{2} \cos(\omega_c - \omega_m) t \tag{4}$$

Sidebands: In equation (4) we have three terms :

i. $E_c \cos \omega_c t$, represents original carrier voltage.

ii. $\frac{m_a E_c}{2} \cos(\omega_c + \omega_m) t$, represents upper sideband

iii. $\frac{m_a E_c}{2} \cos(\omega_c - \omega_m) t$, represents the lower sideband.

The lower sideband term with angular frequency $(\omega_c - \omega_m)$ and upper sideband term with angular frequency $(\omega_c + \omega_m)$ are located on either side of the carrier at a frequency interval of ω_m in their frequency spectrum as shown in fig. 1. The magnitude of the amplitude of both the upper and lower sidebands is $\frac{m_a E_c}{2}$. If m_a is unity, then each sideband term is half the carrier voltage in amplitude. The carrier voltage component transmits no information, while each of the two sidebands carries complete intelligence.

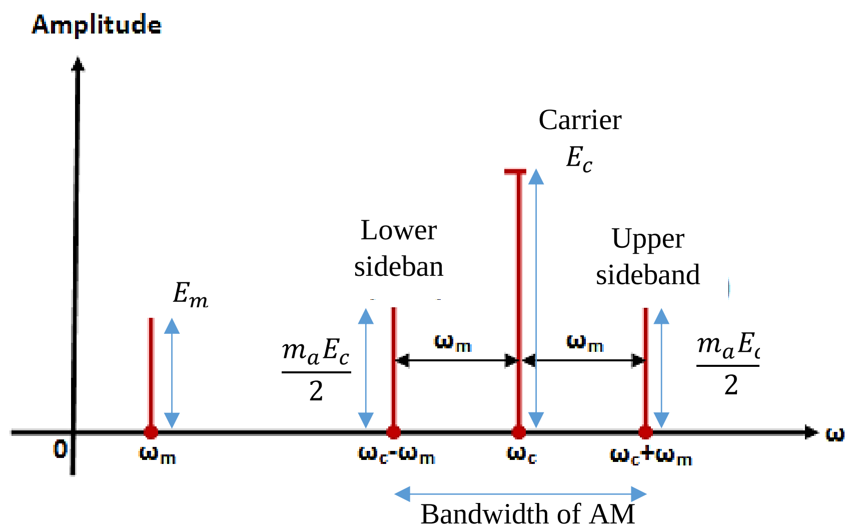


Fig. 1: Plot of frequency spectrum of amplitude modulated

Waveforms: Figure (2) shows the waveforms of modulating signal and that of a modulated wave called the modulation envelope. It is clear from fig. 2 (b) that

$$m_a E_c = E_{max} - E_c \tag{5a}$$

$$m_a = \frac{E_{max} - E_c}{E_c}$$

Also

$$m_a E_c = E_c - E_{min} \tag{5b}$$

$$m_a = \frac{E_c - E_{min}}{E_c}$$

Equating both, (5a) and (5b)

$$E_{max} - E_c = E_c - E_{min}$$

$$2E_c = E_{max} + E_{min}$$

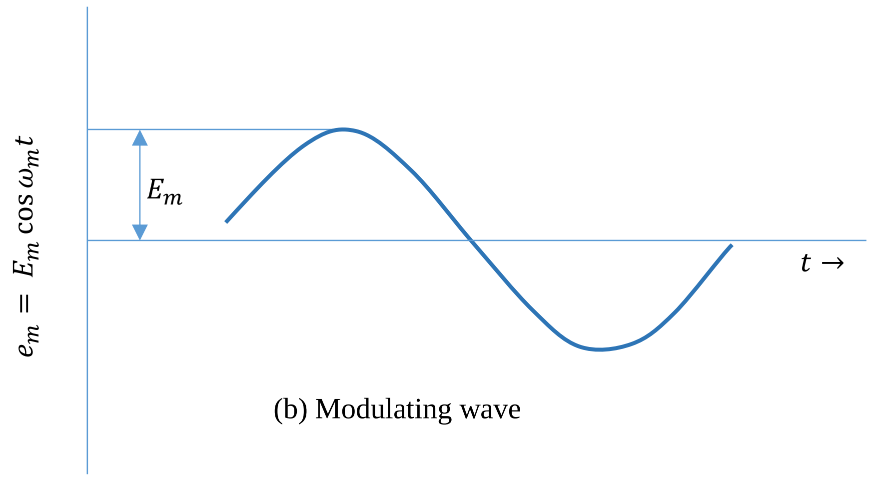
Putting this value of E_c in 5(a), we get

$$m_a = \frac{E_{max} - \frac{E_{max}}{2} - \frac{E_{min}}{2}}{\frac{E_{max} + E_{min}}{2}}$$

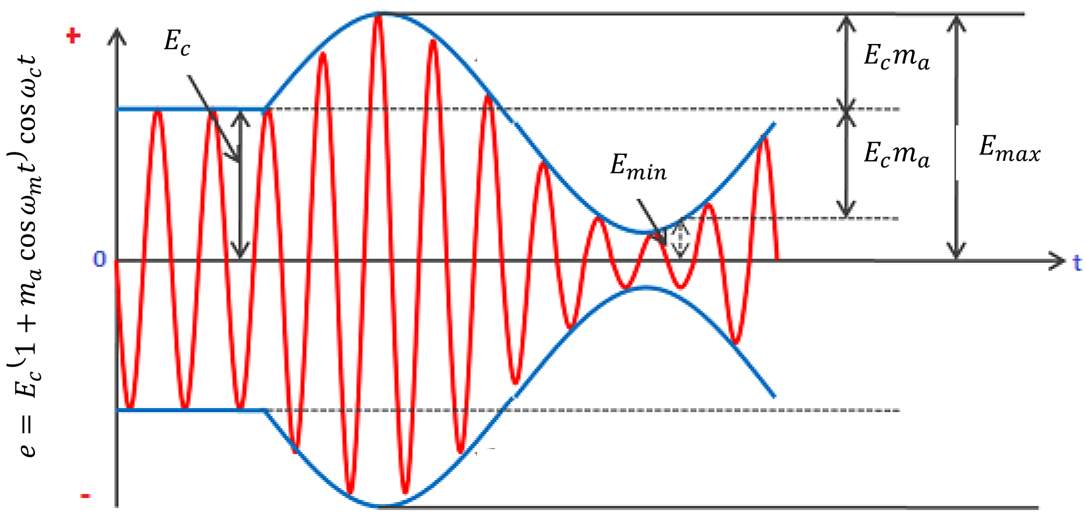
$$= \frac{E_{max} - E_{min}}{E_{max} + E_{min}} \tag{5c}$$

And therefore percentage of modulation

$$m_a = \frac{E_{max} - E_{min}}{E_{max} + E_{min}} \times 100\ \% \tag{5d}$$



(b) Modulating wave



(a) Amplitude modulation

Fig. 2: Amplitude-modulated wave representation.

Modulated waves with, various degrees of modulation: Modulated wave is expressed as

$$e = E_c(1 + m_a \cos \omega_m t) \cos \omega_c t$$

where m_a is modulation factor or modulation index, and its value ranges from zero to one. We shall now take the following different cases:

(i) **When $m_a = 0$**: which means no amplitude modulation is present and the expression for e reduces

to

$$e = E_m \cos \omega_m t$$

representing the pure carrier.

(ii) **When $m_a = 0.5$**: which means the percentage of modulation is 50%. In other words, the extent to which the carrier is being modulated is equal to half of its amplitude.

(iii) **When $m_a = 1$**: which implies 100% modulation. In other words, the extent to which the carrier is modulated is equal to its amplitude. This means that at the negative peaks of the modulating signal, carrier amplitude is reduced to zero.

(iv) **When $m_a > 1$**: represents the case of **overmodulation**. For such a case expression for e is no longer valid. At the negative peaks of modulating signal, no modulated wave exists and thus such a case results in a distortion of the intelligence that is being transmitted.

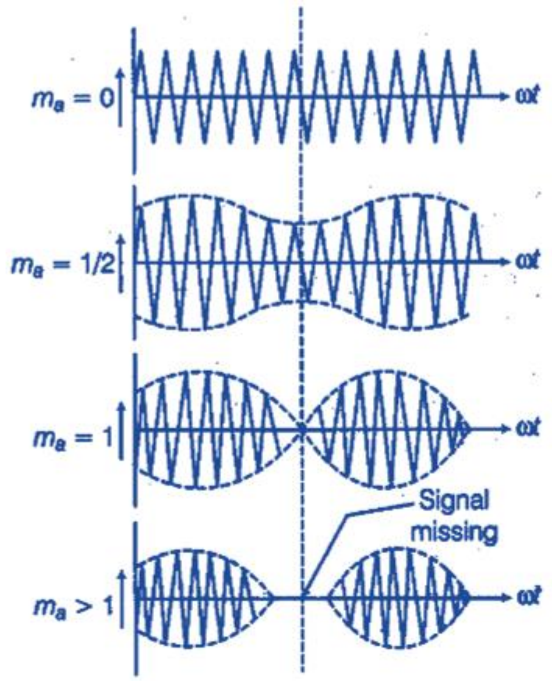


Fig. 3: Modulated wave with various degrees of modulation.

Energy Distribution Among the Three Components of Amplitude Modulated Wave

The voltage components contained in the wave, represented by the equation

$$e = E_c \cos \omega_c t + \frac{m_a E_c}{2} \cos(\omega_c + \omega_m)t + \frac{m_a E_c}{2} \cos(\omega_c - \omega_m)t$$

have r.m.s. values equal to $\frac{E_c}{\sqrt{2}}$, $\frac{m_a E_c}{2\sqrt{2}}$ and $\frac{m_a E_c}{2\sqrt{2}}$ respectively. Therefore, when such a modulated wave, having these three components, is impressed upon some resistance R, then the total average power dissipated would be equal to

$$P_{av} = \frac{{E_c}^2}{2R} + \frac{{m_a}^2 {E_c}^2}{8R} + \frac{{m_a}^2 {E_c}^2}{8R}$$

$$= \frac{{E_c}^2}{2R}\left(1 + \frac{{m_a}^2}{2}\right)$$

Where $\frac{{E_c}^2}{2R}$ represents the power dissipated by the carrier components alone. Obviously,

$$P_{av} > \frac{{E_c}^2}{2R} > P_c$$

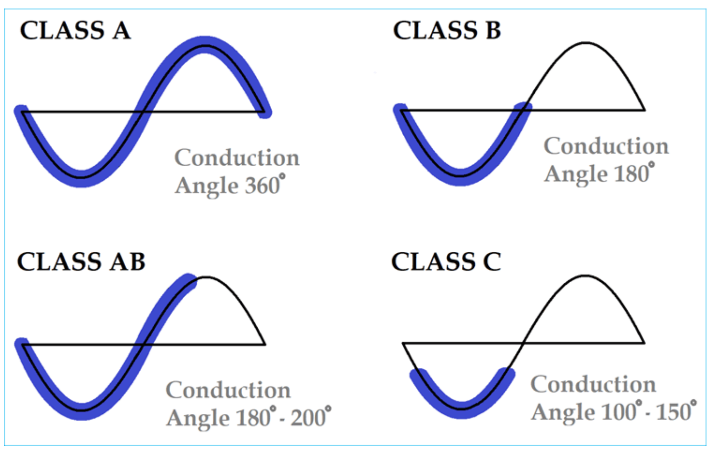
That is, power content of modulated wave exceeds that of the corresponding unmodulated carrier. This additional power supplied by modulating source, in modulated wave furnishes the energy contained in side bands of the modulated wave. It is the useful energy that carries intelligence and its relative magnitude is proportional to ${m_a}^2$. The ratio of useful power contained in side bands to the total power contained in the modulated wave is

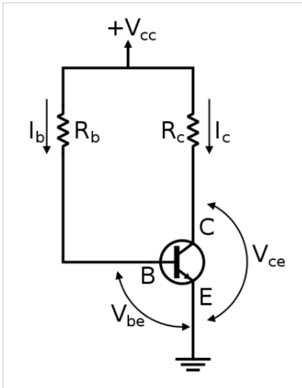
$$\frac{P_{side\ bands}}{P_{av}} = \frac{2\frac{{m_a}^2{E_c}^2}{8R}}{\frac{{E_c}^2}{2R}\left(1+\frac{{m_a}^2}{2}\right)}$$

$$= \frac{\frac{{m_a}^2}{2}}{1+\frac{{m_a}^2}{2}}$$

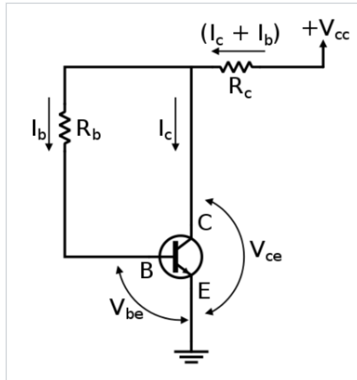
Suppose modulation is 100% then $m_a = 1$. For such a case, $\frac{P_{side\ bands}}{P_{av}} = \frac{\frac{1}{2}}{1+\frac{1}{2}} = \frac{1}{3}$.

That is, only one-third of the total power of the modulated wave is contained in the two side bands. The rest two-third resides in the carrier component which is of no use. If in the modulated wave, the carrier be suppressed then the modulator would draw less power from the supply source and the process of modulation will be quite economical.

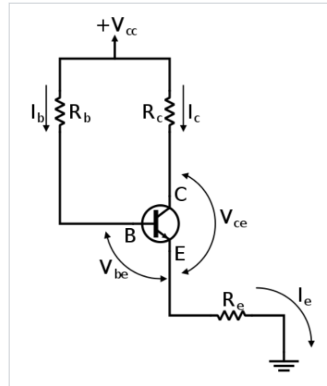




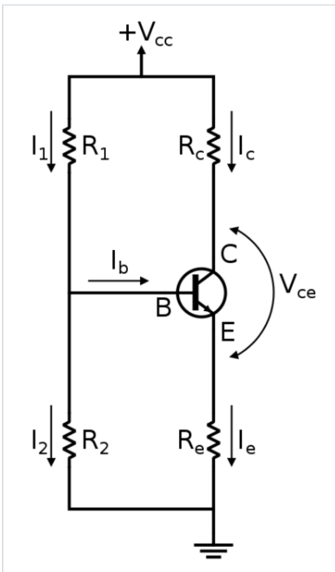
Fixed bias (Base bias)



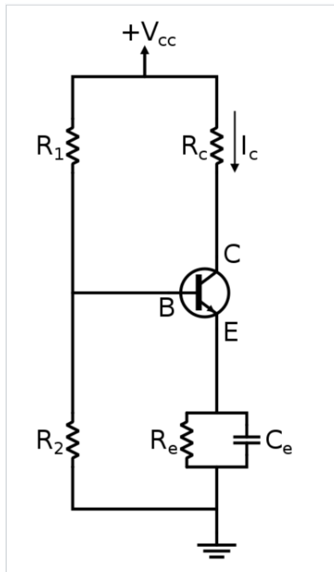
Collector-to-base bias



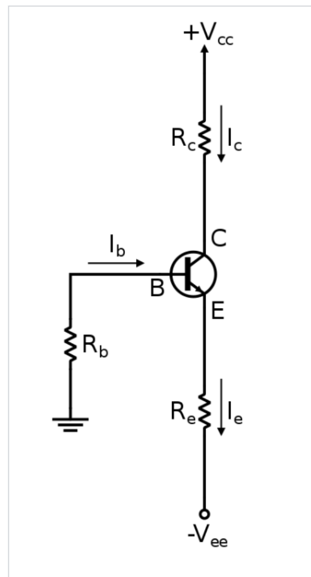
Fixed bias with emitter resistor



Voltage divider bias



Voltage divider with capacitor



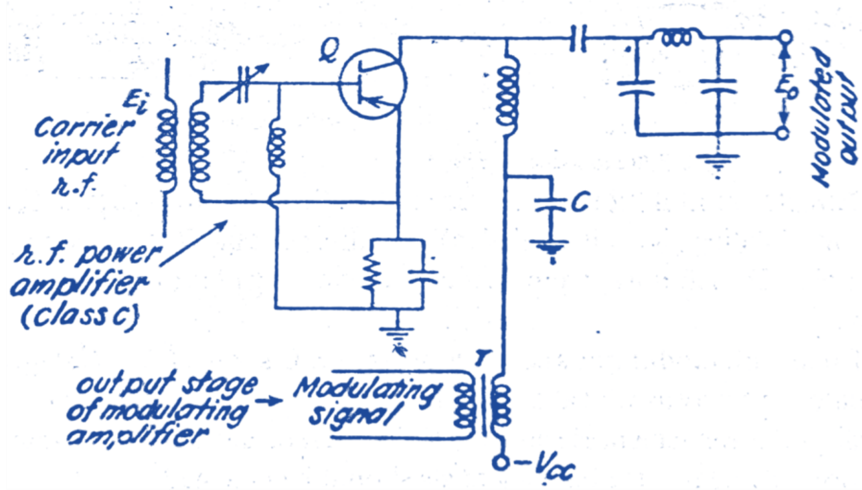
Emitter bias

Transistor Modulators

Transistor *RF* power amplifiers can be employed for the purpose of modulation in a way similar to vacuum tube circuits. Modulating signal can be fed to any of the three elements viz. collector modulation, base modulation and emitter modulation circuits.

Collector Modulation

In the fig, modulating signal has been applied to the collector, in series with $d.c.$ collector supply voltage. Transformer, T, matches the output stage of modulating amplifier with the collector load of class-C $r.f.$ power amplifier Q. Capacitor, C, offers low reactance to radio frequencies but a high reactance to modulating signal frequencies. Thus, it grounds the carrier and serves to keep ref. out of modulating circuit and out of power supply. To obtain 100% modulation, maximum value of modulating voltage, E_{mm}, must equal power supply voltage, V_{CC}. Under these conditions, the $r.f.$ output of the modulated amplifier is equal to zero at the negative peak of the modulating signal. Then we have



$$E_{mm} = V_{CC}$$

Since the average collector current of the modulated amplifier is reduced to zero from I_C during the negative half cycle of the modulated signal, the maximum value of modulating signal current is

$$I_{mm} = I_C$$

Thus, modulator power output will be

$$P_0 = \frac{E_{mm} I_{mm}}{2} = \frac{V_{CC}\, I_C}{2}$$

$$= \frac{1}{2} \times \textit{the power supplied by the power supply}$$

The effective load on the modulator is given by

$$R_L = \frac{E_{mm}}{I_{mm}} = \frac{V_{CC}}{I_C}$$

Collector modulation has the advantages over base modulation (to be described) of better linearity, higher collector efficiency, and higher power output per transistor, but also a disadvantage of requiring more modulating power.

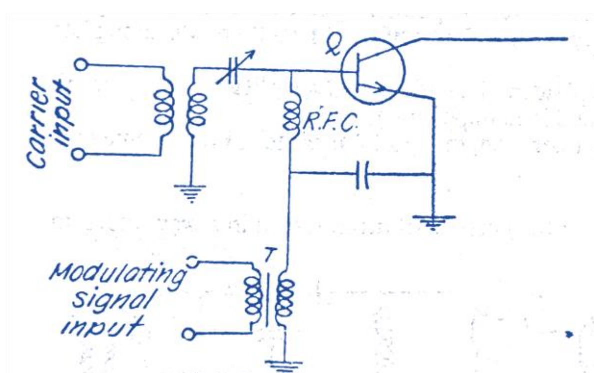
Base Modulation

The modulating signal is fed into the base circuit of the transistor. Its salient features are:

(i) it has a low modulating power requirement compared to collector modulation.

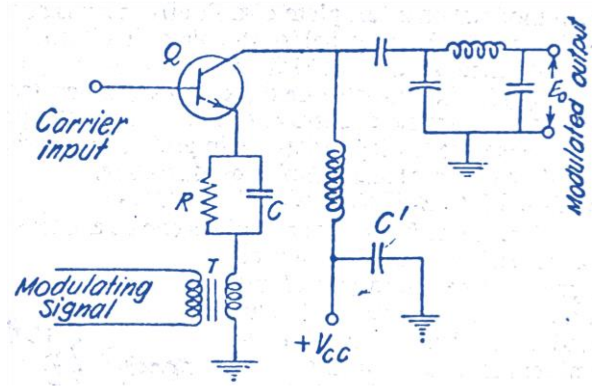
(ii) it has poorer linearity and is more difficult to adjust, and

(iii) power output and efficiency are also low.



Emitter modulation

Its characteristics lie in between base and collector modulation. The modulating signal is applied to the emitter circuit. Resistor R, in combination with C, develops the proper bias. Transformer, T, provides better matching of the impedance. C' serves to keep $r.f.$ out of power supply.



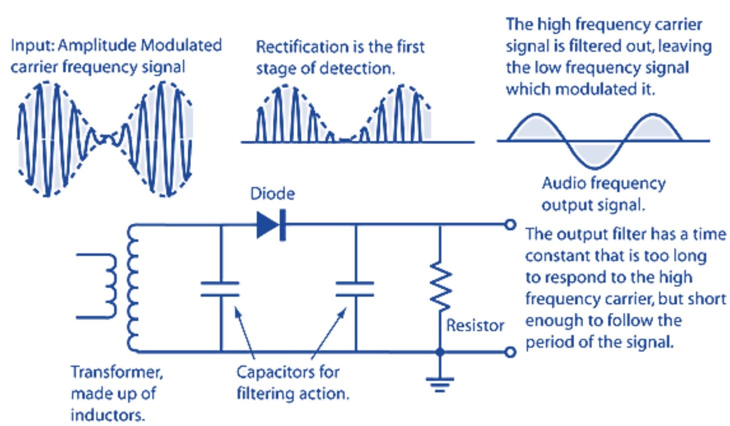
Demodulation

Demodulation is extracting the original information-bearing signal from a carrier wave. A demodulator is an electronic circuit (or computer program in a software-defined radio) that is used to recover the information content from the modulated carrier wave. There are many types of modulation so there are many types of demodulators. The signal output from a demodulator may represent sound (an analog audio signal), images (an analog video signal) or binary data (a digital signal).

These terms are traditionally used in connection with radio receivers, but many other systems use many kinds of demodulators. For example, in a modem, which is a contraction of the terms modulator/demodulator, a demodulator is used to extract a serial digital data stream from a carrier signal which is used to carry it through a telephone line, coaxial cable, or optical fiber.

AM Demodulation

The modulated AM carrier wave is received by the antenna of the radio receiver and is rectified by the action of a detector diode. Then the rectified signal passes through a low-pass filter for which the time constant is too long to respond to the high frequency of the AM carrier wave. AM carriers are in the range 600 to 1400 kHz. The signal frequency which modulates it is much lower, 0.02 to 5 kHz, and it can pass through the filter.



The AM detector shown here would be suitable for picking up only one AM radio station for which the values of the capacitors and inductors were chosen. A practical AM radio use the process called heterodyning to shift the carrier frequency of each radio station to a single frequency called an Intermediate Frequency or IF so that a single sophisticated AM detection circuit can be used to receive all AM radio stations.

FM Demodulation

The first demodulation technique that we'll look at begins with a high-pass filter. We'll assume that we're dealing with narrowband FM. We need to design the high-pass filter such that the attenuation will vary significantly within a frequency band whose width is twice the bandwidth of the baseband signal. Let's explore this concept more thoroughly.

The received FM signal will have a spectrum that is centered around the carrier frequency. The width of the spectrum is approximately equal to twice the bandwidth of the baseband signal; the factor of two results from the shifting of the positive and negative baseband frequencies (as discussed here), and it is "approximately" equal because the integration applied to the baseband signal can affect the shape of the modulated spectrum. Thus, the lowest frequency in the modulated signal is approximately equal to the carrier frequency minus the highest frequency in the baseband signal, and the highest frequency in the modulated signal is approximately equal to the carrier frequency plus the highest frequency in the baseband signal.

Our high-pass filter needs to have a frequency response that causes the lowest frequency in the modulated signal to be attenuated significantly more than the highest frequency in the modulated signal. If we apply this filter to an FM waveform, what will be the result? It will be something like shown in fig. 2. This plot shows both the original FM waveform and the high-pass-filtered waveform, for purposes of comparison. Fig.3 shows just the filtered waveform, so that you can see it more clearly.

By applying the filter, we have turned frequency modulation into amplitude modulation. This is a convenient approach to FM demodulation, because it allows us to benefit from envelope-detector circuitry that has been developed for use with amplitude modulation. **The filter used to produce this waveform was nothing more than an RC high-pass with a cutoff frequency approximately equal to the carrier frequency.**

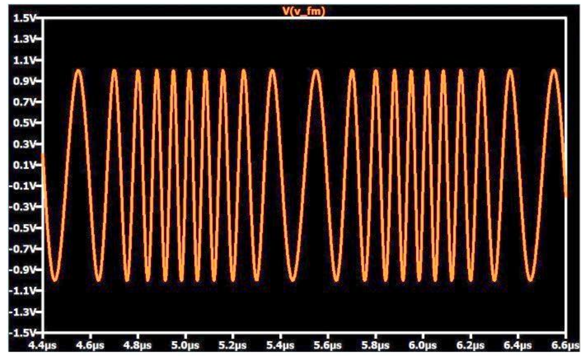


Fig. 1: Frequency modulated signal.

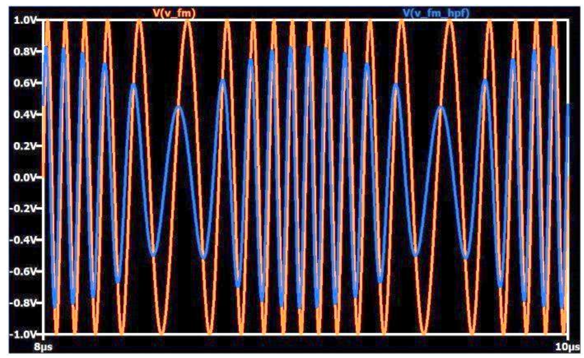


Fig. 2: FM signal after passing high pass filter.

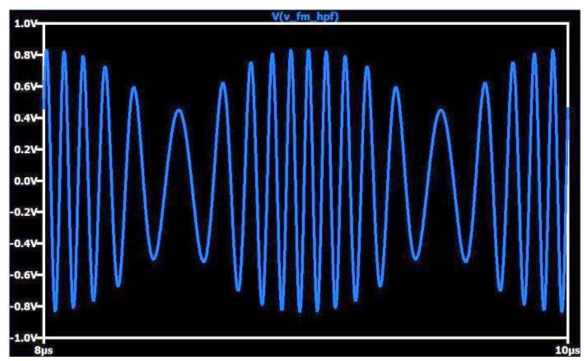
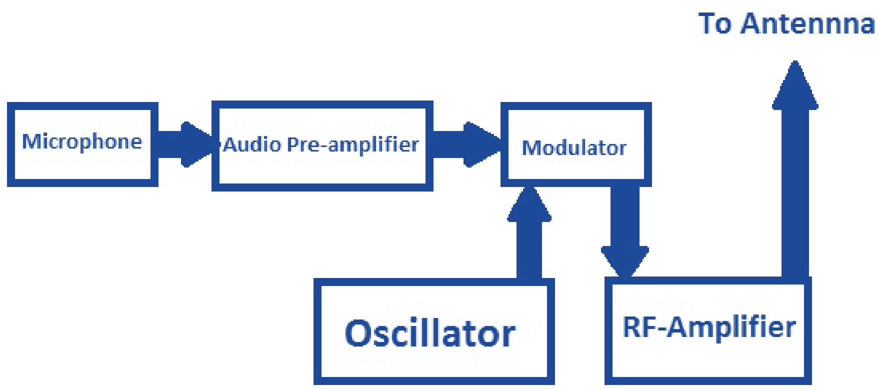


Fig. 3: Filtered waveform.

Transmitter Circuits

In electronics and telecommunications, a radio transmitter or just transmitter is an electronic device which produces radio waves with an antenna. The transmitter itself generates a radio frequency alternating current, which is applied to the antenna. When excited by this alternating current, the antenna radiates radio waves.

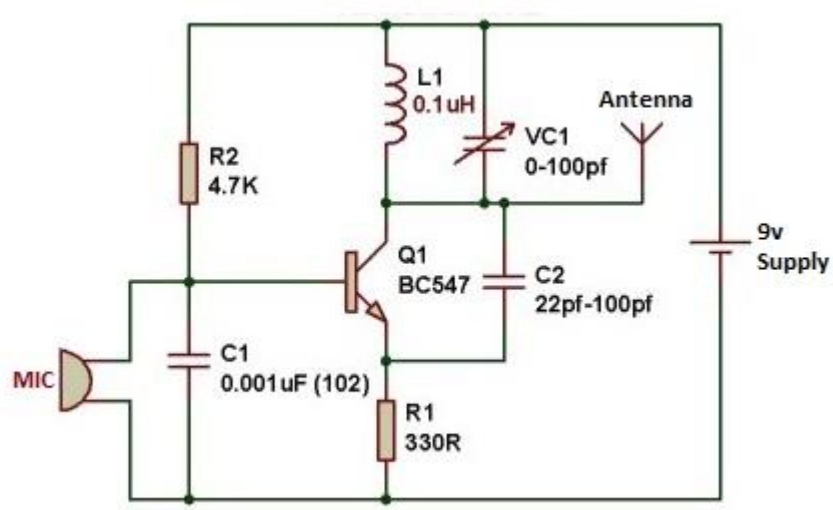
Transmitters are necessary component parts of all electronic devices that communicate by radio, such as radio and television broadcasting stations, cell phones, walkie-talkies, wireless computer networks, Bluetooth enabled devices, garage door openers, two-way radios in aircraft, ships, spacecraft, radar sets and navigational beacons. The term transmitter is usually limited to equipment that generates radio waves for communication purposes; or radiolocation, such as radar and navigational transmitters. Generators of radio waves for heating or industrial purposes, such as microwave ovens or diathermy equipment, are not usually called transmitters, even though they often have similar circuits.



Block diagram of FM transmitter

FM Transmitter

The circuit diagram shown for FM transmitter circuit and the electronic components are the resistor, capacitor, trimmer or variable capacitor, inductor (coil), transmitter, mic, 9v of power supply and an antenna. The mic or a microphone is assumed to grab the sound signals and there is a presence of sensor with capacitance value inside the mic. The change in pressure of air or AC signal causes in production of such capacitance.



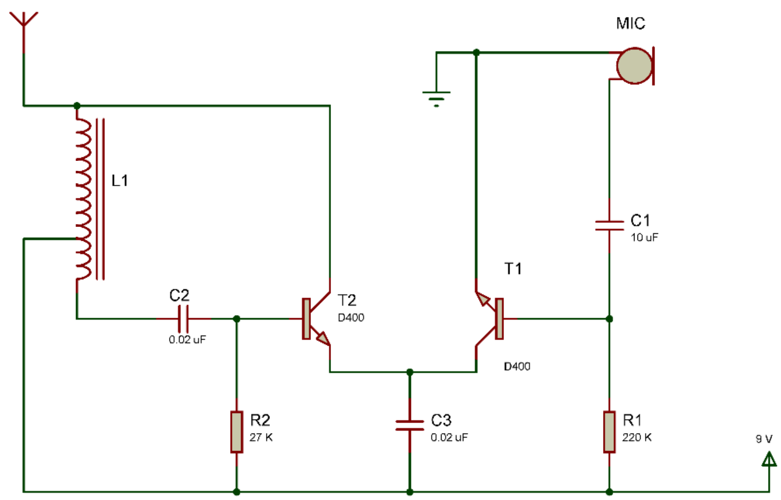
The Oscillation circuit could be made with the help of Transistor BC547 or 2N3904, Inductor and the variable capacitor. The Transistor BC547 or 2N3904 used in FM Transmitter Circuit. It is a NPN transistor which is basically used for signal and voltage amplification purposes. If the current is passed through L1 Inductor and the variable capacitor, the FM Transmitter Circuit will start oscillating with the resonant to that of the carrier frequency (I.e., frequency of the carrier signal).

The negative current or the negative feedback will result C2 Capacitor to the Transmitter Circuit. The oscillator is required in FM Transmitter Circuit in order to generate the radio frequency carrier waves. The transmitter circuit is capable of storing the energy for the oscillation as it is a derived from the (inductor and capacitor) LC circuits. The audio i.e., input signal obtained from the mic is passed through the base of the transistor in order to modulate the output signal of the LC circuit in FM (i.e., frequency modulation wave) Form.

Here, the main purpose of the variable capacitor becomes that to vary the resonant frequency in order to obtain the best FM Signal Frequency band. The modulated signal is then transmitted or radiated as a radio wave with the frequency of FM frequency range. The antenna is nothing but a piece of good conductor, in our application we'd used a copper wire of 30cm long and 26 gauges thick. You can use copper wire up to 25-27 inches long as an antenna in the circuit, the length of the antenna however must be significant.

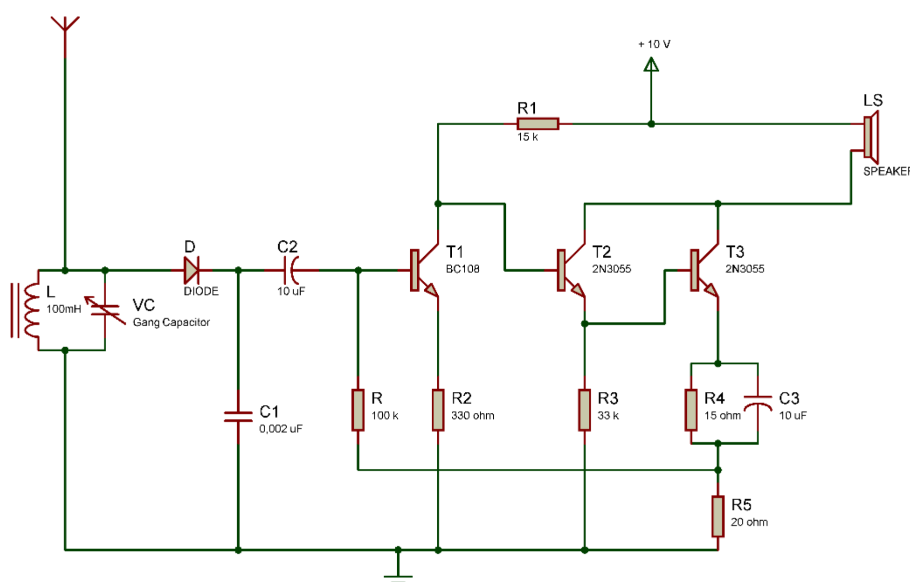
AM Transmitter

The radio transmitter is an electronic device, which transmits the modulated carrier waves through the antenna. The audio signal from a microphone is mixed by the oscillator transistor with the carrier wave produced by the Hartley oscillator. The modulated wave is then amplified and radiated through the antenna. A typical transmitter circuit is shown in the fig. the first section consists of a transducer (dynamic microphone) and a capacitor that converts the audio human voice (elastic wave in the air) to an electrical signal. This signal is then amplified by a pre-amplifier using the transistor T1. The amplified signal is then applied to the Hartley oscillator consisting of a transistor T2 and a tuning circuit with L and C2 (gang capacitor). Finally, the modulated signal from the oscillator section is radiated through the antenna as an electromagnetic wave.



Am Radio Receiver

A radio receiver is a device that can receive the radio frequency energy transmitted from a transmitting station and then demodulates or separates the audio signal from the high-frequency carrier wave. Usually, the audio signal and the carrier signal are modulated and sent it in space through transmitter antenna as electromagnetic (e.m.) waves. These e.m. waves induce fluctuating voltage (of the order of a few microvolts) to the receiver antenna. Which are demodulated by the semiconductor diode detector D and the capacitor C1. The demodulated audio signal is amplified by the voltage amplifier T1. The transistors



T2 and T3 act as power amplifiers. The R4 and C3 combination provides self-biasing of the transistor T3 for A-class amplification. Automatic gain control (AGC) is done by the resistors R and R5. The amplified electrical pulses are then converted to an audible signal by the loudspeaker (LS).

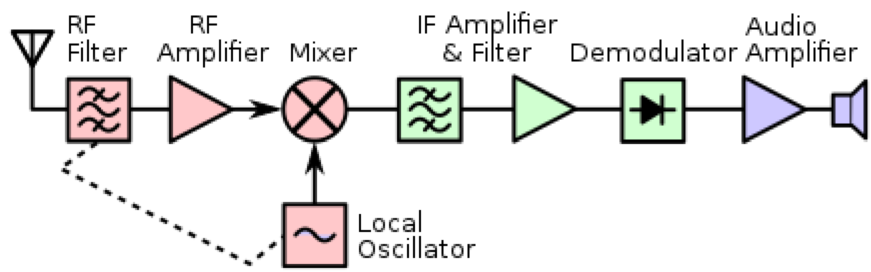
Heterodyne principle

The principle that multiple frequencies applied to a nonlinear device produces new frequencies that are sums and differences of the applied frequencies and their harmonics. Heterodyning, or mixing, is the process of multiplying a weak signal by a strong sinusoidal carrier, sometimes called the local oscillator, to shift the frequency of the signal in such a way that the information carried by the signal is preserved.

Superheterodyne Receiver

A superheterodyne receiver, often shortened to superhet, is a type of radio receiver that uses frequency mixing to convert a received signal to a fixed intermediate frequency (IF) which can be more conveniently processed than the original carrier frequency. It was long believed to have been invented by US engineer Edwin Armstrong, but after some controversy, the earliest patent for the invention is now credited to French radio engineer and the radio manufacturer Lucien Lèvy. Virtually all modern radio receivers use the superheterodyne principle.

The diagram at the right shows the block diagram of a typical single-conversion superheterodyne receiver. The diagram has blocks that are common to superheterodyne receivers, with only the RF amplifier being optional.

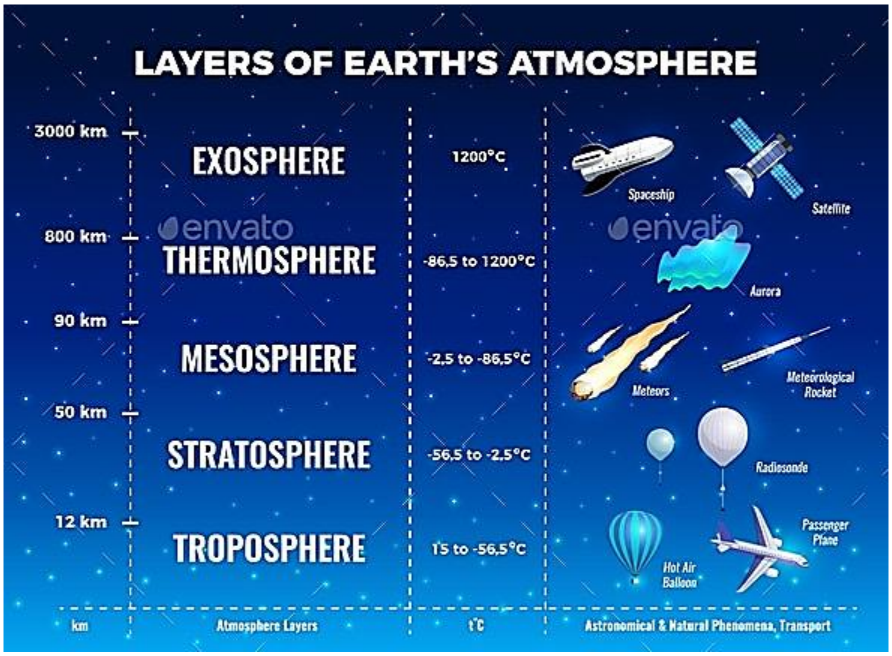


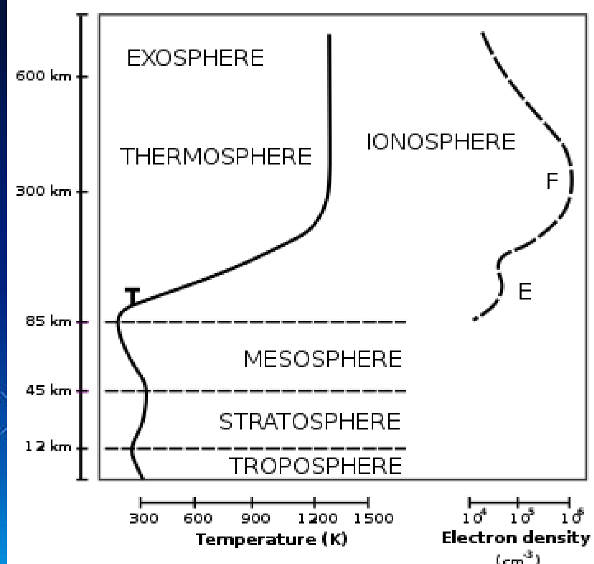
Block diagram of a typical superheterodyne receiver. Red parts are those that handle the incoming radio frequency (RF) signal; green are parts that operate at the intermediate frequency (IF), while blue parts operate at the modulation (audio) frequency. The dotted line indicates that the local oscillator and RF filter must be tuned in tandem

The antenna collects the radio signal. The tuned RF stage with an optional RF amplifier provides some initial selectivity; it is necessary to suppress the image frequency, and may also serve to prevent strong out-of-passband signals from saturating the initial amplifier. A local oscillator provides the mixing frequency; it is usually a variable frequency oscillator which is used to tune the receiver to different stations. The frequency mixer does the actual heterodyning that gives the superheterodyne its name; it changes the incoming radio frequency signal to a higher or lower, fixed, intermediate frequency (IF). The IF band-pass filter and amplifier supply most of the gain and the narrowband filtering for the radio. The demodulator extracts the audio or other modulation from the IF radiofrequency. The extracted signal is then amplified by the audio amplifier.

Circuit description

To receive a radio signal, a suitable antenna is required. The output of the antenna may be very small, often only a few microvolts. The signal from the antenna is tuned and may be amplified in a so-called radio frequency (RF) amplifier, although this stage is often omitted. One or more tuned circuits at this stage block frequencies that are far removed from the intended reception frequency. To tune the receiver to a particular station, the frequency of the local oscillator is controlled by the tuning knob (for instance). Tuning of the local oscillator and the RF stage may use a variable capacitor, or varicap diode. The tuning of one (or more) tuned circuits in the RF stage must track the tuning of the local oscillator.





Radio Wave Propagation

To understand the meaning of ground wave propagation, it is important to know the meaning of radio wave propagation. Radio wave propagation is the behavior of radio waves as they propagate from one point to another or into various parts of the atmosphere.

These waves propagation can be classified depending upon the frequencies as:

- Ground waves propagation
- Skywave propagation
- Free space propagation

Ground Wave Propagation

Ground wave propagation is a type of radio propagation which is also known as a surface wave. These waves propagate over the earth's surface in low and medium frequencies. These are mainly used for transmission between the surface of the earth and the ionosphere. These are made up of the number of constituent waves.

The reason why it is known as a ground wave is that it is the sum of the waves that are reflected by the earth's surface or any hills. The waves follow the curvature of the earth, enabling them to cover beyond the horizon. Beyond the horizon, the waves get blocked by the curvature of the earth and the signals are produced by the diffracted surface wave.

Following is the table explaining the frequency of ground waves depending on the type of ground:

Advantages of Ground Wave Propagation

- These waves have the tendency to bend around the corners or obstructions during propagation which makes them more efficient and also these are not affected by the change in atmospheric conditions.

Disadvantages of Ground Wave Propagation

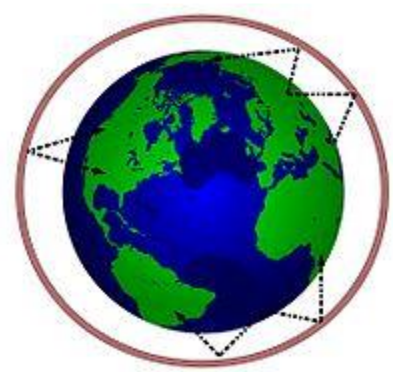
- High-frequency waves cannot be transmitted as the energy losses are more because of the absorption of energy in the earth's atmosphere.
- These are used to cover short ranges and also involve attenuation of waves as they interact with the eddy currents produced by the surface of the earth.

Applications Ground Wave Propagation

- These can be used for one-way communication from the military to submerged submarines as they penetrate to a significant depth into seawater.
- AM, FM, and television broadcasting can be done with the help of ground waves.

Skywave Propagation (Propagation through Ionosphere)

The ionosphere is a region of the upper atmosphere, from about 80 km to 1000 km in altitude, where neutral air is ionized by solar photons and cosmic rays. When high-frequency signals enter the ionosphere at a low angle they are bent back towards the earth by the ionized layer. If the peak ionization is strong enough for the chosen frequency, a wave will exit the bottom of the layer earthwards – as if obliquely reflected from a mirror. Earth's surface (ground or water) then reflects the descending wave back up again towards the ionosphere.



When operating at frequencies just below the MUF, losses can be quite small, so the radio signal may effectively "bounce" or "skip" between the earth and ionosphere two or more times (multi-hop propagation), even following the curvature of the earth. Consequently, even signals of only a few Watts can sometimes be received many thousands of miles away. This is what enables shortwave broadcasts to travel all over the world. If the ionization is not great enough, the wave only curves slightly downwards, and subsequently upwards as the ionization peak is passed so that it exits the top of the layer only slightly displaced. The wave then is lost in space. To prevent this, a lower frequency must be chosen. With a single "hop", path distances up to 3500 km may be reached. Longer transmissions can occur with two or more hops.

Skywave propagation also known as the skip is a type of radio wave propagation. It is either the reflected or refracted back waves to the earth from the ionosphere which is an electrically charged layer of the upper atmosphere.

Medium and shortwave frequencies can be refracted back to earth which is beyond the horizon which makes them useful in the transcontinental transmission of the waves. Following is the table explaining the distribution of frequencies MUF ($maximum\ usable\ frequency$) and OWF ($optimum\ working\ frequency$) depending upon the layers of the earth:

Layers	D	E	F1
MUF (MHz)	16	28	16
OWF (MHz)	13.6	23.8	13.6

Critical Frequency

Critical frequency is defined as the maximum frequency at which the total internal reflection takes place from the ionosphere. The mathematical representation is given as:

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

Where,
f_c: critical frequency in Hz
N_{max}: maximum electron density per m3

The mathematical representation of critical frequency as a function of MUF is:

$$f_c = \frac{MUF}{\sec\theta}$$

Where,
f_c: critical frequency in Hz
MUF: maximum usable frequency
θ: angle of incidence

Critical frequency varies depending upon atmospheric conditions, time of the day, and the angle of fire of the radio waves by the antenna.

Skip Distance

Skip distance is defined as the minimum distance from the earth's surface and the point from where the radio signal is been transmitted. For a flat earth skip distance is given as:

$$D_{skip} = 2h\sqrt{\left(\frac{MUF}{f_c}\right)^2 - 1}$$

Where,
D_{skip}: skip distance
h: height at which reflection happens
MUF: maximum usable frequency
f_c: critical frequency

Skywave propagation applications:

- Satellite communications take place with the help of skywave propagation as it is dependent on the upper atmospheric conditions
- Mobile communications

Space Wave Propagation

Space wave propagation is defined for the radio waves that occur within the $20\,km$ of the atmosphere i.e.; troposphere, comprising of direct and reflected waves. These waves are also known as tropospheric propagation as they can travel directly from the earth's surface to the troposphere surface of the earth. It is also known as line-of-sight propagation as the signals are sent in a straight line from the transmitter to the receiver.

In order to prevent attenuation and loss of signal strength, the height of the antennas and distance between them can be given as:

$$D_m = (2RH_t)^{-\frac{1}{2}} + (2RH_r)^{-\frac{1}{2}}$$

Where,
D_m: distance between the two antennas
R: radius of the earth
H_t: height of transmission antenna
H_r: height of receiver antenna

Applications of space wave propagation

It is used in various communication systems like

- A line of sight communication and satellite communication
- Radar communication
- Microwave linking

Space wave propagation limitations

- These waves are affected by the curvature of the earth.
- The propagation of these waves happens along with the line-of-sight distance which is defined as the distance between the transmitting antenna and the receiving antenna which is also known as the range of communication.

Antenna Fundamentals

Antennas are widely used in the field of telecommunications and we know many applications for them. Antennas receive an electromagnetic wave and convert it to an electric signal, or receive an electric signal and radiate it as an electromagnetic wave. In this article, we are going to look at the science behind antennas.

Difference between fluctuating and radiating electromagnetic fields:

We have an electric signal, so how do we convert it to an electromagnetic wave? You might have a simple answer in your mind. That is to use a closed conductor, and with the help of the principle of electromagnetic induction, you will be able to produce a fluctuating magnetic field and an electric field around it as shown in Fig:1A. However, this fluctuating field around the source is of no use in transmitting signals. The electromagnetic field here does not propagate; instead, it just fluctuates around the source. In an antenna, the electromagnetic waves need to be separated from the source and they should propagate (Fig:1B). Before looking at how an antenna is made, let's understand the physics behind the wave separation.

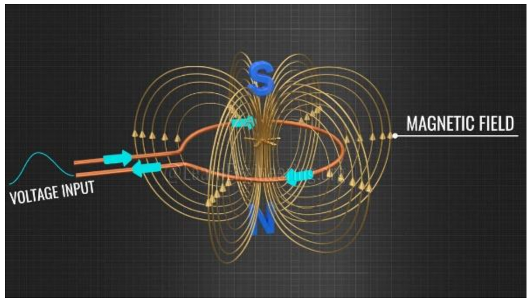


Fig:1A: Fluctuating electromagnetic field in electromagnetic induction

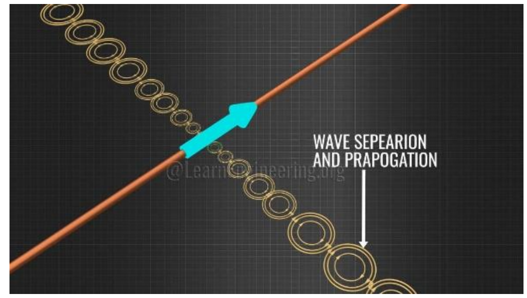


Fig:1B: Radiating electromagnetic field in a hypothetical antenna

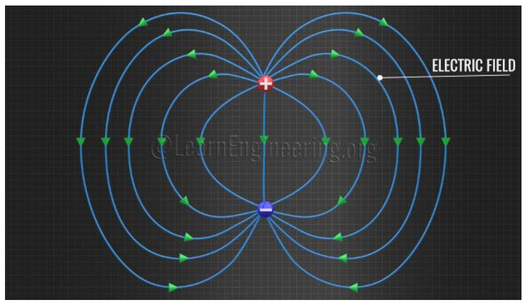


Fig:2A: The electric field lines of an electric dipole are extended from the positive to the negative charge

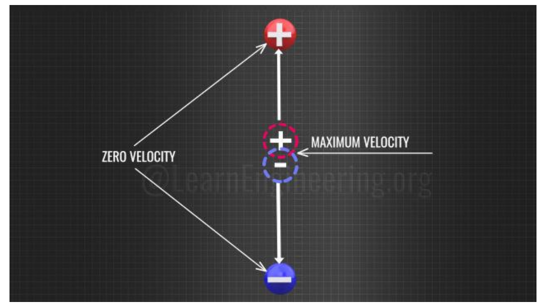


Fig:2B: Acceleration and deceleration of charge particles

Physics behind the oscillating dipole and radiation:

Consider one positive and one negative charge placed a distance apart. This arrangement is known as a dipole, and they obviously produce an electric field as shown in Fig:2A. Now, assume that these charges are oscillating as shown in Fig:2B. At the midpoint of their path, the velocity will be at the maximum, and at the ends of their paths the velocity will be zero. The charged particles undergo continuous acceleration and deceleration due to this velocity variation.

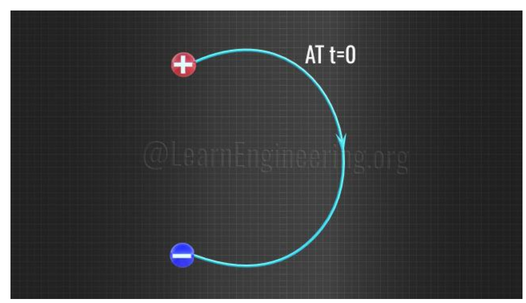


Fig:3: Electric field is shown at t=0

1. Electric field line at t=0

The challenge now is to find out how the electric field varies due to this movement. Let's concentrate on only one electric field line (Fig:3).

2. Electric field line at t=T/8

The wavefront formed at time zero expands and is deformed as shown after one-eighth of a time period (Fig:4A). This is surprising; you might have expected a simple electric field as shown at this location. Why has the electric field stretched and formed a field like this? as shown in Fig:4B This is because the accelerating or decelerating charges produce an electric field with some memory effects. The old electric field does not easily adjust to the new

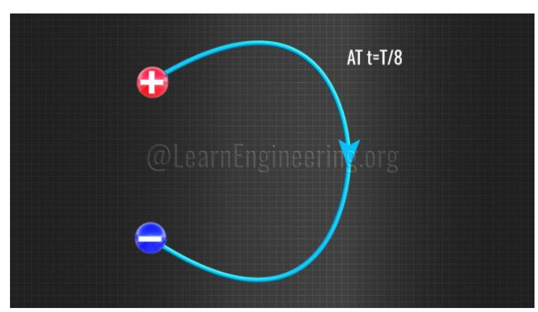


Fig:4A: At t=T/8, the expected shape of the electric field condition. We need to spend some time to understand this memory effect of the electric field, or kink generation, of accelerating or decelerating charges.

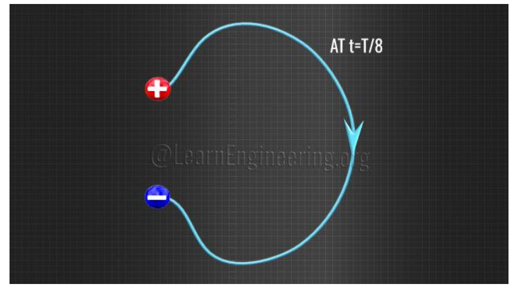


Fig:4B: At t=T/8, the actual shape of the electric field

3. Electric field line at t=T/4

If we continue our analysis in the same manner, we can see that at one-quarter of a time period, the wavefront ends meet at a single point (Fig:5).

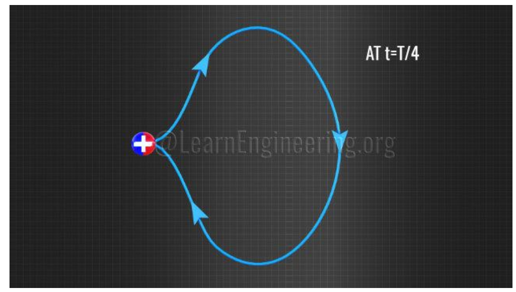


Fig:5: At t=T/4, the ends of electric field meets at a single point and the separation and propagation happens

After this, the separation and propagation of the wavefront happen. If you draw electric field intensity variation with the distance, you can see that the wave propagation is sinusoidal in nature (Fig:6). It is interesting to note that the wavelength of the propagation so produced is exactly double that of the length of the dipole. We will come back to this point later. Please note that this varying electric field will automatically generate a varying magnetic field perpendicular to it. This is exactly what we need in an antenna. In short, we can make an antenna, if we can make an arrangement for oscillating the positive and negative charges.

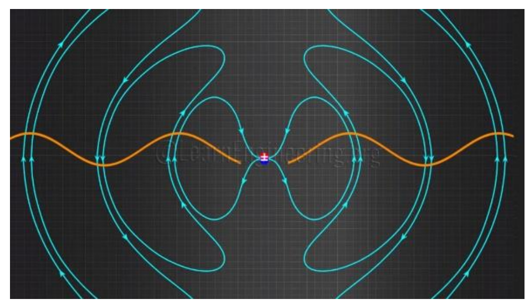


Fig:6 Electromagnetic radiation in a dipole

How does radiation happen in antennas?

In practice, the production of such an oscillating charge is very easy. Take a conducting rod with a bend in its center, and apply a voltage signal at the center (7A). Assume this is the signal you have applied, a time-varying voltage signal. Consider the case at time zero. Due to the effect of the voltage, the electrons will be displaced from the right of the dipole and will be accumulated on the left. This means the other end, which has lost electrons, automatically becomes positively charged (7B). This arrangement has created the same effect as the previous dipole charge case, i.e. positive and negative charges at the end of a wire. With the variation of voltage with time, the positive and negative charges will shuttle to and fro.

The simple dipole antenna also produces the same phenomenon we saw in the previous section and wave propagation occurs. We have now seen how the antenna works as a transmitter. The frequency of the transmitted signal will be the same as the frequency of the applied voltage signal. Since the propagation travels at the speed of light, we can easily calculate the wavelength of the propagation (Fig:8). For the perfect transmission, the length of the antenna should be half of the wavelength.

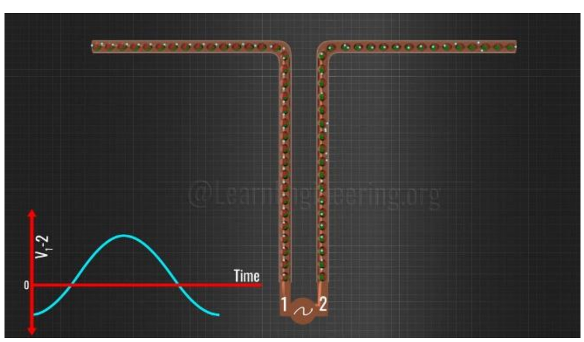


Fig:7A A long straight wire with an AC source at its center is a dipole antenna radiating electromagnetic waves

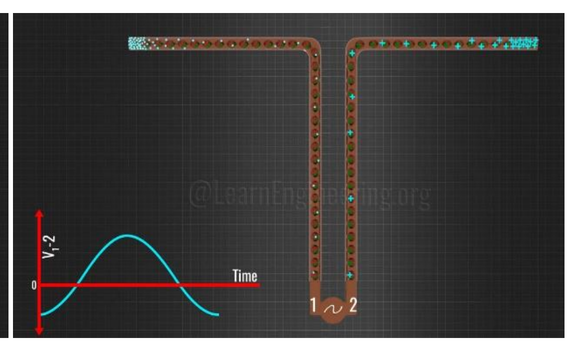


Fig:7B When a time varying voltage signal is applied, the electrons are accumulated at one end and created positive charges on other end

$$f_{antenna} = f_{input}$$

$$C = f_{antenna} \times \lambda_{antenna}$$

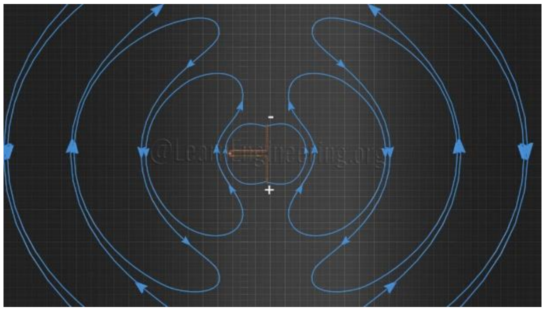


Fig:8 Antenna radiate electromagnetic waves at the speed of light

How do antennas receive signals?

The operation of the antenna is reversible and it can work as a receiver if a propagating electromagnetic field hits it. Let's see this phenomenon in detail.

Take the same antenna again and apply an electric field. At this instant, the electrons will accumulate at one end of the rod. This is the same as an electric dipole. As the applied electric field varies, the positive and negative charges accumulate at the other ends. The varying charge accumulation means a varying electric voltage signal is produced at the center of the antenna. This voltage signal is the output when the antenna works as a receiver as shown in Fig:9. The frequency of the output voltage signal is the same as the frequency of the receiving EM wave. It is clear from the electric field configuration that for perfect reception, the size of the antenna should be half of the wavelength. In all these discussions we have seen that the antenna is an open circuit.

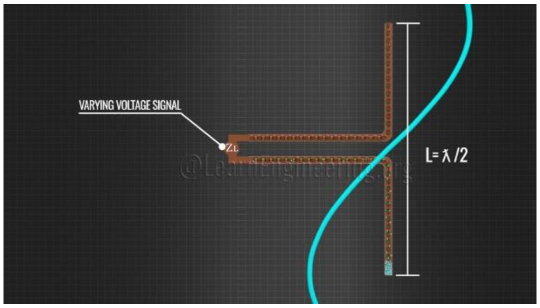


Fig:9 An Antenna can work as a receiver if a propagating electromagnetic field hits it

Construction and Working of Yagi Uda Antenna

In the past, dipole antennas were used for TV reception. The colored bar acts as a dipole and receives the signal as shown in the figure. The dipole is the main driven element of it. A reflector and director are also needed in this kind of antenna to focus the signal on the dipole. The reflector element is always longer and the director element are always shorter than the driven element. This complete structure is known as a Yagi-Uda antenna (Fig:1). The yagi uda antenna was invented by two Japanese scientists Hidetsugu Yagi and Shintaro Uda. It is a directional antenna and is used in point-to-point communication. The driven element or dipole antenna converted the received signal into electrical signals and these electrical signals were carried by coaxial cable to the television unit(Fig:2).

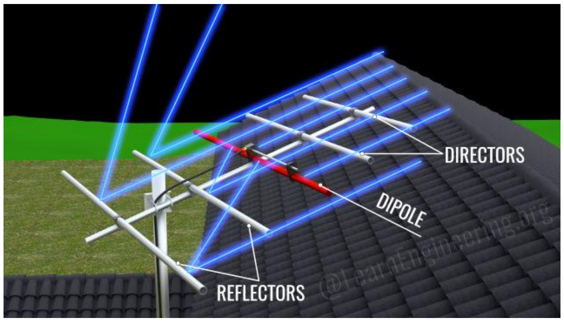


Fig:1: A Yagi uda antenna consist of dipole, directors and reflectors

Fig:2: The yagi uda antenna converts the received signals into electrical signals and these signals are carried by coaxial cable to the television unit

Construction and Working of Satellite Dish Antenna

Nowadays we have moved to dish TV antennas. These consist of two main components, a parabolic-shaped reflector and a low noise block down converter. The parabolic dish receives electromagnetic signals from the satellite and focuses them onto the LNBF as shown in Fig:1. The shape of the parabolic is very specifically and accurately designed.

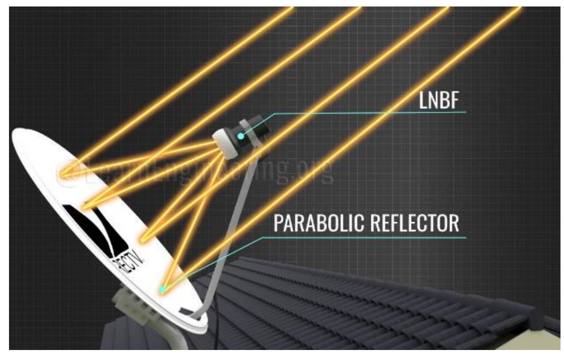


Fig:1: In satellite dish antenna the incoming signal focused on to LNBF via

The LNBF is made up of a feed horn, a waveguide, a PCB, and a probe (2A). The incoming signals are focused onto the probe via the feed horn and waveguide. At the probe, voltage is induced as we saw in the simple dipole case. The voltage signal so generated is fed to a PCB for signal processing such as filtration, conversion from high to low frequency, and amplification. After signal processing, these electrical signals are carried down to the television unit through a coaxial cable (Fig:2B).

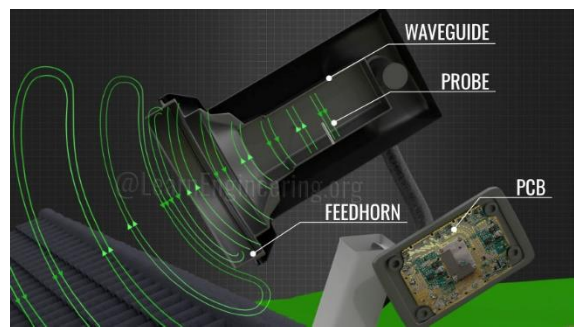


Fig:2A: A detailed structure of Low Noise Blockdown Feedhorn (LNBF)

If you open up an LNB you will most probably find 2 probes instead of one, the second probe being perpendicular to the first one. The 2-probe arrangement means the available spectrum can be used twice, by sending the waves with either horizontal or vertical polarization. One probe detects the horizontally polarized signal and the other the vertically polarized signal as shown in Fig:3.

Fig:2B The satellite dish converts the received electromagnetic signals into electrical signals which are carried out by coaxial cable to the television unit

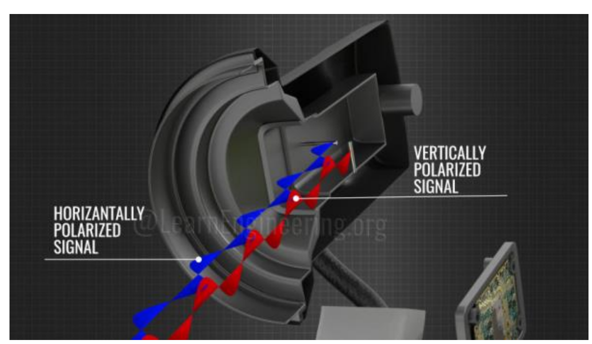


Fig:3: The horizontal and vertical probe detects horizontally polarized signal and vertically polarized signal respectively.

Construction and Working of Microstrip antenna or Patch antenna

The cellphone in your hand uses a completely different type of antenna, called a patch antenna (Fig:1A). These types of antennas are inexpensive and fabricated easily onto a printed circuit board. A patch antenna consists of a metallic patch or strip placed on a ground plane with a piece of dielectric material in-between. Here, the metallic patch acts as a radiating element. The length of the metal patch should be half of the wavelength for proper transmission and reception (Fig:1B). Please note that the description of the patch antenna we explained here is very basic.

Fig:1A: Planar inverted F antenna, a type of patch antenna used in modern cell phones.

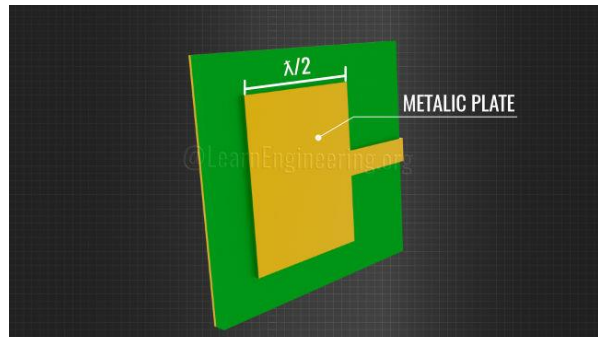
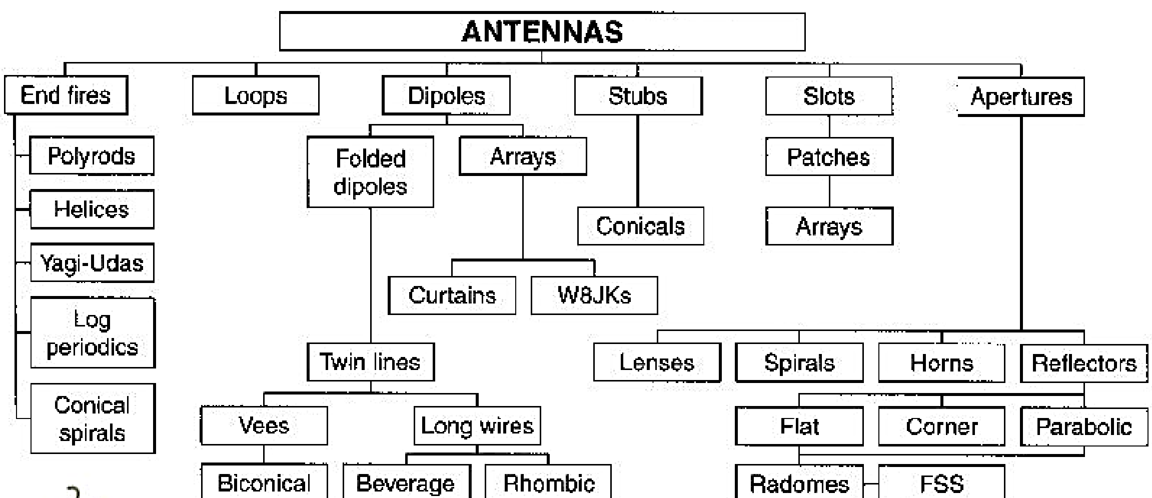
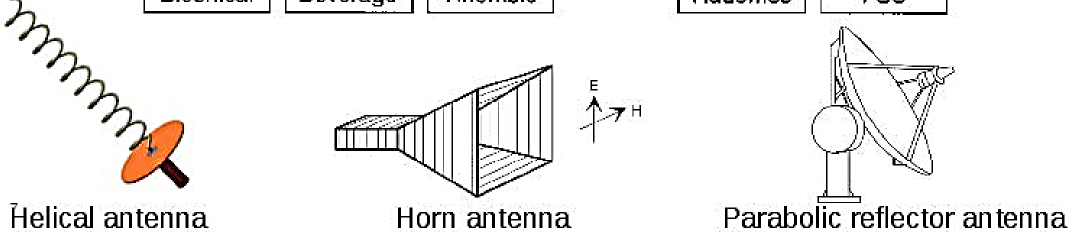


Fig:1B: A schematic of simple patch antenna

Types of Antenna and Arrays

In radio systems, many different antenna types are used with specialized properties for particular applications. Antennas can be classified in various ways. The list below groups together antennas under common operating principles:





References

- Hand Book of Electronics – Gupta & Kumar
- Internet
- https://www.youtube.com/watch?v=ZaXm6wau-jc
- https://en.wikipedia.org/wiki/Antenna_types
- https://www.lesics.com/how-does-an-antenna-work.html