Problem Solver (DI03000121) - Problem Solver

Antigravity AI

June 4, 2026

Contents

CHAPTER 1

Basics of Communication System

1.1 Introduction

Communication is the process of exchanging information between a source and a destination. Electronic communication systems facilitate this process using electrical signals or electromagnetic waves. This unit covers the fundamental concepts of communication systems, including the electromagnetic spectrum, signal representation, block diagrams, modulation, and noise.

1.2 Electromagnetic (EM) Wave Spectrum

The electromagnetic (EM) wave spectrum encompasses all types of electromagnetic radiation, ranging from extremely low-frequency radio waves to high-frequency gamma rays. In communication systems, specific frequency bands are allocated for various applications to prevent interference and ensure efficient transmission.

1.2.1 Frequency Bands and Applications

The radio frequency (RF) spectrum is divided into several bands:

- **Extremely Low Frequency (ELF)** (3 Hz - 30 Hz): Submarine communication.
- **Very Low Frequency (VLF)** (3 kHz - 30 kHz): Navigation systems.
- **Low Frequency (LF)** (30 kHz - 300 kHz): AM broadcasting (long wave) and navigational beacons.
- **Medium Frequency (MF)** (300 kHz - 3 MHz): Standard AM radio broadcasting.
- **High Frequency (HF)** (3 MHz - 30 MHz): Shortwave radio and amateur radio.
- **Very High Frequency (VHF)** (30 MHz - 300 MHz): FM radio broadcasting and television channels.
- **Ultra High Frequency (UHF)** (300 MHz - 3 GHz): Mobile phones, Wi-Fi, and television broadcasting.
- **Super High Frequency (SHF)** (3 GHz - 30 GHz): Satellite communication and radar.
- **Extremely High Frequency (EHF)** (30 GHz - 300 GHz): Radio astronomy and millimeter-wave communication.

1.3 Signals and Their Representation

A signal is a physical quantity that varies with time space or any other independent variable and carries information.

1.3.1 Analog and Digital Signals

- **Analog Signal:** A continuous-time signal where the varying quantity can take any value in a continuous range. For example, human speech converted into an electrical signal using a microphone.
- **Digital Signal:** A discrete-time signal that takes on only discrete values. Typically, digital signals are represented as binary sequences (0s and 1s).

1.3.2 Standard Test Signals

- **Impulse Signal:** A signal that has infinite amplitude at time $t = 0$ and zero elsewhere, with an area of one.
- **Pulse Signal:** A signal that remains at a constant amplitude for a specific duration and is zero otherwise.
- **Sinusoidal Signal:** A continuous periodic signal representing smooth oscillations, mathematically expressed as $x(t) = A\sin(\omega t + \phi)$.
- **Rectangular Signal:** A periodic or non-periodic signal that alternates between two fixed amplitude levels instantaneously.
- **Saw-tooth Signal:** A signal that increases linearly with time and then rapidly drops to its starting value.

1.3.3 Time and Frequency Domain Representation

- **Time Domain:** Represents how the amplitude of the signal changes with respect to time.
- **Frequency Domain:** Represents the signal's spectral components, showing the amplitude and phase of each frequency present in the signal. Fourier transform is used to convert signals from the time domain to the frequency domain.

1.4 Block Diagram of Communication System

1.4.1 Analog Communication System

The basic components of an analog communication system include:

- **Information Source:** Generates the message signal (e.g., voice, video).
- **Transducer:** Converts physical quantities into electrical signals.
- **Transmitter:** Processes the electrical signal for transmission, typically involving modulation and amplification.
- **Communication Channel:** The medium through which the signal travels (e.g., coaxial cable, optical fiber, free space). It introduces noise and distortion.
- **Receiver:** Captures the transmitted signal and extracts the original message signal through demodulation.
- **Destination Transducer:** Converts the electrical signal back into its original physical form (e.g., a loudspeaker).

1.4.2 Digital Communication System

A digital communication system consists of additional blocks for digitizing the signal:

- **Source Encoder:** Converts the digital signal into a more efficient representation to reduce redundancy.
- **Channel Encoder:** Adds redundant bits to the data for error detection and correction.
- **Digital Modulator:** Maps the digital bits into analog waveforms suitable for transmission over the channel.

- **Digital Demodulator:** Reconstructs the digital sequence from the received waveforms.
- **Channel Decoder:** Detects and corrects errors introduced by the channel.
- **Source Decoder:** Reconstructs the original digital signal from the compressed data.

1.5 Modulation

Modulation is the process of varying one or more properties of a periodic waveform, called the carrier signal, with a modulating signal that contains information to be transmitted.

1.5.1 Need for Modulation

- **Antenna Height Reduction:** Low-frequency baseband signals require impractically large antennas. Modulation shifts the signal to a higher frequency, reducing the required antenna size.
- **Multiplexing:** Multiple signals can be transmitted over a single channel simultaneously using different carrier frequencies.
- **Noise Immunity:** Modulated signals (especially FM and digital modulation) are less susceptible to noise and interference.

1.5.2 Classification of Modulation

Modulation techniques are primarily classified based on the type of carrier signal:

1. **Continuous Wave (Analog) Modulation:** The carrier is a continuous sinusoidal wave.
    - **Amplitude Modulation (AM):** The amplitude of the carrier wave is varied in proportion to the modulating signal.
    - **Frequency Modulation (FM):** The frequency of the carrier wave is varied in proportion to the modulating signal.
    - **Phase Modulation (PM):** The phase of the carrier wave is varied in proportion to the modulating signal.
2. **Pulse Modulation:** The carrier is a train of rectangular pulses.
    - **Analog Pulse Modulation:** Pulse Amplitude Modulation (PAM), Pulse Width Modulation (PWM), and Pulse Position Modulation (PPM).
    - **Digital Pulse Modulation:** Pulse Code Modulation (PCM) and Delta Modulation (DM).

1.6 Noise in Communication System

Noise is an unwanted electrical signal that interferes with the transmission and reception of the desired message signal. It degrades the quality of communication and limits the system's performance.

1.6.1 Classification of Noise

Noise can be broadly classified into two main categories:

1. **External Noise:** Originates outside the communication system.
    - **Atmospheric Noise:** Caused by natural electrical disturbances like lightning.
    - **Extraterrestrial Noise:** Includes solar noise and cosmic noise from outer space.
    - **Industrial Noise:** Man-made noise from electrical motors, automobile ignition systems, and power lines.

2. **Internal Noise:** Generated within the communication equipment itself.
   - **Thermal Noise (Johnson Noise):** Caused by the random motion of electrons in a conductor due to thermal agitation.
   - **Shot Noise:** Arises due to the discrete nature of charge carriers in electronic devices like diodes and transistors.
   - **Flicker Noise:** Low-frequency noise associated with surface defects in semiconductor devices.

1.6.2 Signal-to-Noise Ratio and Noise Figure

Methodology:

Calculating SNR and Noise Figure **Signal-to-Noise Ratio (SNR)** is the ratio of signal power to noise power. It is usually expressed in decibels (dB).

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{1.1}$$

$$SNR_{dB} = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{1.2}$$

Noise Figure (NF) is a measure of degradation of the SNR as the signal passes through a system. It is defined as the ratio of the input SNR to the output SNR.

$$F = \frac{SNR_{in}}{SNR_{out}} \tag{1.3}$$

In decibels, it is expressed as:

$$NF_{dB} = 10\log_{10}(F) = SNR_{in,dB} - SNR_{out,dB} \tag{1.4}$$

Worked Example:

SNR Calculation A communication system receives a signal power of 10ţW and the noise power measured at the same point is $10nW$. Calculate the Signal-to-Noise Ratio (SNR) in linear scale and in decibels.
Solution: Given: Signal Power, $P_s =$ 10ţ$W = 10 \times 10^{-6}W$ Noise Power, $P_n = 10nW = 10 \times 10^{-9}W$
Step 1: Calculate linear SNR.

$$SNR = \frac{P_s}{P_n} = \frac{10 \times 10^{-6}}{10 \times 10^{-9}} = 1000$$

Step 2: Calculate SNR in decibels (dB).

$$SNR_{dB} = 10\log_{10}(SNR)$$

$$SNR_{dB} = 10\log_{10}(1000) = 10 \times 3 = 30 \text{ dB}$$

Worked Example:

Noise Figure Calculation An amplifier has an input signal-to-noise ratio of 40 dB and an output signal-to-noise ratio of 32 dB. Find the Noise Figure (NF) of the amplifier.
Solution: Given: Input SNR, $SNR_{in,dB} = 40$ dB Output SNR, $SNR_{out,dB} = 32$ dB
The Noise Figure in dB is given by the difference between the input SNR and the output SNR in dB:

$$NF_{dB} = SNR_{in,dB} - SNR_{out,dB}$$

$$NF_{dB} = 40 \text{ dB} - 32 \text{ dB} = 8 \text{ dB}$$

CHAPTER 2

Amplitude Modulation (AM) Communication

2.1 Introduction

Communication systems often require the transmission of low-frequency signals (such as audio or data) over long distances. To achieve this efficiently, the low-frequency baseband signal is superimposed onto a high-frequency carrier wave. This process is known as modulation. In this chapter, we will explore Amplitude Modulation (AM), one of the fundamental techniques used in electronic communication.

2.2 Need of Modulation

Transmitting baseband signals directly over a communication channel presents several practical challenges. Modulation is essential for the following reasons:

- **Antenna Height:** For efficient transmission and reception, the length of the antenna should be at least one-quarter of the wavelength ($\lambda/4$) of the signal. For low-frequency audio signals (e.g., 15 kHz), the required antenna height is impracticably large (about 5 km). By modulating a high-frequency carrier, the required antenna size is significantly reduced.
- **Multiplexing:** Modulation allows multiple signals from different sources to be transmitted simultaneously over the same channel without interference. By using different carrier frequencies for different signals, frequency division multiplexing (FDM) can be achieved.
- **Reduction in Noise and Interference:** Modulating a signal onto a high-frequency carrier can help in reducing the effects of external noise and interference during transmission.
- **Overcoming Equipment Limitations:** High-frequency signals are less prone to attenuation in certain transmission media, allowing communication over greater distances.

2.3 Types of Modulation Techniques

Modulation techniques are broadly classified based on the type of carrier wave used. The two main categories are Continuous Wave (CW) Modulation and Pulse Modulation.

2.3.1 Continuous Wave (CW) Modulation

In CW modulation, the carrier is a continuous high-frequency sinusoidal wave. Based on which parameter of the carrier wave is varied, CW modulation is further divided into:

- **Amplitude Modulation (AM):** The amplitude of the carrier wave is varied in accordance with the instantaneous amplitude of the modulating signal.
- **Frequency Modulation (FM):** The frequency of the carrier wave is varied in accordance with the instantaneous amplitude of the modulating signal.
- **Phase Modulation (PM):** The phase of the carrier wave is varied in accordance with the instantaneous amplitude of the modulating signal.

2.3.2 Pulse Modulation

In pulse modulation, the carrier is a periodic train of pulses. The parameters of the pulses (such as amplitude, width, or position) are varied according to the modulating signal.

2.4 Amplitude Modulation (AM)

In Amplitude Modulation (AM), the maximum amplitude of the high-frequency carrier wave is varied in proportion to the instantaneous amplitude of the low-frequency modulating signal, while keeping the frequency and phase of the carrier constant.

2.4.1 Signals in AM

- **Modulating Signal (Baseband Signal):** This is the original information signal that contains the data to be transmitted. It is a low-frequency signal, denoted as $m(t) = A_m \sin(\omega_m t)$.
- **Carrier Signal:** This is the high-frequency sinusoidal wave that carries the modulating signal. It is denoted as $c(t) = A_c \sin(\omega_c t)$, where $\omega_c \gg \omega_m$.

2.4.2 Modulation Index

The modulation index (or depth of modulation) is a measure of the extent to which the carrier amplitude is varied by the modulating signal. It is defined as the ratio of the maximum amplitude of the modulating signal to the maximum amplitude of the unmodulated carrier.

$$m = \frac{A_m}{A_c} \tag{2.1}$$

Where A_m is the peak amplitude of the modulating signal and A_c is the peak amplitude of the carrier. The modulation index is often expressed as a percentage (% modulation $= m \times 100$). Based on the value of m, modulation can be classified as:

- **Under Modulation ($m < 1$):** The AM wave envelope does not reach zero. This is the ideal condition for transmission without distortion.
- **Critical Modulation ($m = 1$):** The AM wave envelope just touches zero.
- **Over Modulation ($m > 1$):** The AM wave envelope crosses zero, leading to envelope distortion (clipping). This causes severe distortion at the receiver end.

Methodology:

Calculating Modulation Index from Waveform If the maximum ($V_{\max}$) and minimum ($V_{\min}$) amplitudes of the AM wave envelope are known, the modulation index can be calculated as:

$$m = \frac{V_{\max} - V_{\min}}{V_{\max} + V_{\min}} \tag{2.2}$$

Worked Example:

Modulation Index Calculation An AM wave has a maximum peak-to-peak voltage of 10 V and a minimum peak-to-peak voltage of 2 V. Calculate the modulation index.
Solution: Given: $V_{\max} = 10/2 = 5$ V (since peak-to-peak is given) $V_{\min} = 2/2 = 1$ V
Using the formula:

$$m = \frac{V_{\max} - V_{\min}}{V_{\max} + V_{\min}}$$

$$m = \frac{5-1}{5+1} = \frac{4}{6} = 0.67$$

The modulation index is 0.67 or 67%.

2.4.3 Mathematical Representation of AM Signal

Let the modulating signal be $m(t) = A_m \sin(\omega_m t)$ and the carrier signal be $c(t) = A_c \sin(\omega_c t)$. The instantaneous voltage of the AM wave is given by:

$$v_{AM}(t) = [A_c + m(t)] \sin(\omega_c t) \tag{2.3}$$

Substituting $m(t)$:

$$v_{AM}(t) = [A_c + A_m \sin(\omega_m t)] \sin(\omega_c t) \tag{2.4}$$

Taking A_c common:

$$v_{AM}(t) = A_c \left[1 + \frac{A_m}{A_c} \sin(\omega_m t)\right] \sin(\omega_c t) \tag{2.5}$$

Since $m = \frac{A_m}{A_c}$:

$$v_{AM}(t) = A_c[1 + m \sin(\omega_m t)] \sin(\omega_c t) \tag{2.6}$$

2.4.4 Representation in Time and Frequency Domain

Expanding the mathematical expression:

$$v_{AM}(t) = A_c \sin(\omega_c t) + mA_c \sin(\omega_m t) \sin(\omega_c t) \tag{2.7}$$

Using the trigonometric identity $2 \sin A \sin B = \cos(A - B) - \cos(A + B)$:

$$v_{AM}(t) = A_c \sin(\omega_c t) + \frac{mA_c}{2} \cos((\omega_c - \omega_m)t) - \frac{mA_c}{2} \cos((\omega_c + \omega_m)t) \tag{2.8}$$

This equation shows that the AM wave consists of three components:

- Unmodulated carrier at frequency f_c with amplitude A_c.
- Lower Sideband (LSB) at frequency $f_c - f_m$ with amplitude $\frac{mA_c}{2}$.
- Upper Sideband (USB) at frequency $f_c + f_m$ with amplitude $\frac{mA_c}{2}$.

2.4.5 Frequency Spectrum and Bandwidth

The frequency spectrum of an AM wave shows these three components. The bandwidth (BW) required for AM transmission is the difference between the highest and lowest frequencies in the spectrum.

$$BW = (f_c + f_m) - (f_c - f_m) = 2f_m \tag{2.9}$$

Thus, the bandwidth of an AM signal is twice the maximum frequency of the modulating signal.

2.4.6 Types of AM Band Spectrum and Applications

- **Double Sideband Full Carrier (DSB-FC):** This is the standard AM signal which contains the carrier and both sidebands. It is used in standard AM radio broadcasting.
- **Double Sideband Suppressed Carrier (DSB-SC):** The carrier is suppressed, and only the two sidebands are transmitted. It saves power but requires a complex receiver. It is used in analog television color information transmission.
- **Single Sideband (SSB):** One sideband and the carrier are suppressed. It saves both power and bandwidth. It is widely used in point-to-point communication, military communication, and amateur radio.
- **Vestigial Sideband (VSB):** One complete sideband and a trace (vestige) of the other sideband are transmitted. It is standard for analog television video transmission.

2.4.7 Power Relations in AM Wave

The total power P_t transmitted in an AM wave is the sum of the carrier power P_c and the power in the sidebands P_{LSB} and P_{USB}.

$$P_t = P_c + P_{LSB} + P_{USB} \tag{2.10}$$

We know that power is proportional to the square of the RMS voltage ($P = V_{rms}^2/R$). Let R be the antenna resistance.

$$P_c = \frac{(A_c/\sqrt{2})^2}{R} = \frac{A_c^2}{2R} \tag{2.11}$$

The amplitude of each sideband is $\frac{mA_c}{2}$.

$$P_{LSB} = P_{USB} = \frac{\left(\frac{mA_c}{2\sqrt{2}}\right)^2}{R} = \frac{m^2 A_c^2}{8R} = \frac{m^2}{4} P_c \tag{2.12}$$

The total sideband power $P_{sb} = P_{LSB} + P_{USB} = \frac{m^2}{2} P_c$. The total power equation becomes:

$$P_t = P_c + \frac{m^2}{2} P_c = P_c \left(1 + \frac{m^2}{2}\right) \tag{2.13}$$

Worked Example:

AM Power Calculation A 400 W carrier is modulated to a depth of 75%. Calculate the total power in the modulated wave.
Solution: Given: Carrier Power (P_c) = 400 W Modulation Index (m) = 0.75
Using the total power formula:

$$P_t = P_c \left(1 + \frac{m^2}{2}\right)$$

$$P_t = 400 \times \left(1 + \frac{0.75^2}{2}\right)$$

$$P_t = 400 \times (1 + 0.28125) = 400 \times 1.28125 = 512.5 \text{ W}$$

The total transmitted power is 512.5 W.

2.4.8 Power Saving in SSB

In a standard DSB-FC AM wave, a large portion of the transmitted power is contained in the carrier, which does not carry any information. The information is solely contained in the sidebands. Total power $P_t = P_c + P_{sb} = P_c(1 + m^2/2)$. If we suppress the carrier and one sideband (SSB transmission), the power saved is significant.

Methodology:

Power Saving Calculation Power saved in SSB = Carrier Power + Power in one sideband

$$\text{Power Saved} = P_c + \frac{m^2 P_c}{4} = P_c \left(1 + \frac{m^2}{4}\right)$$

Percentage power saved $= \frac{\text{Power Saved}}{\text{Total Power}} \times 100$

$$\% \text{ Saving} = \frac{1 + m^2/4}{1 + m^2/2} \times 100$$

For 100% modulation ($m = 1$), the percentage of power saved in SSB transmission compared to standard AM is:

$$\% \text{ Saving} = \frac{1 + 1/4}{1 + 1/2} \times 100 = \frac{1.25}{1.5} \times 100 \approx 83.33\%$$

This shows that SSB transmission is highly power-efficient.

2.5 Generation of AM: DSBSC Signal

A Double Sideband Suppressed Carrier (DSB-SC) signal contains only the upper and lower sidebands without the carrier. It is generated using a balanced modulator circuit.

2.5.1 Balanced Modulator Circuit

A balanced modulator consists of two non-linear devices (like diodes or transistors) connected in a push-pull configuration. The primary function of the balanced modulator is to suppress the carrier signal while allowing the sidebands to pass through.

- The modulating signal $m(t)$ is applied out of phase to the two devices (push-pull input).
- The carrier signal $c(t)$ is applied in phase to both devices.
- The output is taken across a balanced transformer, which subtracts the outputs of the two devices.
- Since the carrier is in phase for both devices, it gets canceled out (suppressed) at the output.
- The sidebands, generated due to the non-linear mixing of the carrier and modulating signal, are out of phase and thus add up at the output, producing the DSB-SC signal.

2.6 Demodulation of AM Signal

Demodulation (or detection) is the process of recovering the original information (modulating signal) from the modulated carrier wave. For AM, the most common and simple demodulator is the envelope detector.

2.6.1 Envelope Detector Using Diode

An envelope detector consists of a diode followed by a parallel RC low-pass filter. It extracts the envelope of the AM wave, which corresponds to the original modulating signal.

- **Rectification:** The diode acts as a half-wave rectifier, allowing only the positive half-cycles of the AM signal to pass.
- **Filtering:** The RC filter removes the high-frequency carrier variations. The capacitor C charges to the peak value of the carrier voltage. When the carrier voltage drops, the capacitor discharges slowly through the resistor R.
- **Time Constant:** The RC time constant is crucial. It must be large enough to hold the charge between carrier cycles, but small enough to follow the variations of the modulating signal.

  $$\frac{1}{f_c} \ll RC \ll \frac{1}{f_m}$$

  If RC is too large, the capacitor discharges too slowly, causing diagonal clipping. If RC is too small, the high-frequency ripple is not adequately filtered.

2.7 Superheterodyne Receiver

Most modern radio receivers use the superheterodyne principle. In a superheterodyne receiver, the incoming high-frequency RF signal is converted to a fixed lower intermediate frequency (IF) before demodulation. This improves selectivity and sensitivity.

2.7.1 Block Diagram and Working

The superheterodyne AM receiver consists of the following blocks:

- **Receiving Antenna:** Intercepts the electromagnetic waves and converts them into small RF voltage signals.
- **RF Amplifier:** Amplifies the weak received RF signal. It also provides some initial selectivity by tuning to the desired frequency band.
- **Local Oscillator:** Generates a high-frequency continuous wave signal (f_{LO}). For AM receivers, the local oscillator frequency is usually higher than the incoming RF frequency (f_{RF}) by the intermediate frequency (f_{IF}). So, $f_{LO} = f_{RF} + f_{IF}$.
- **Mixer:** Mixes the amplified RF signal (f_{RF}) with the local oscillator signal (f_{LO}). The output contains the original frequencies and their sum and difference ($f_{LO} \pm f_{RF}$). The difference frequency is selected as the Intermediate Frequency (IF). For standard AM broadcasting, the IF is commonly 455 kHz.
- **IF Amplifier:** One or more stages of tuned amplifiers fixed at the intermediate frequency (455 kHz). They provide the bulk of the receiver's gain and selectivity.
- **Detector:** An envelope detector (diode detector) that extracts the audio modulating signal from the IF signal.
- **Audio Voltage and Power Amplifiers:** Amplify the weak audio signal to a level sufficient to drive a loudspeaker.
- **Automatic Gain Control (AGC):** A feedback loop that adjusts the gain of the RF and IF amplifiers based on the strength of the received signal, keeping the output volume relatively constant despite variations in signal strength.

2.7.2 Characteristics of a Receiver

The performance of a radio receiver is evaluated based on several characteristics:

- **Sensitivity:** The ability of the receiver to amplify weak signals and produce a usable output. It determines the weakest signal the receiver can successfully pick up.
- **Selectivity:** The ability of the receiver to tune to the desired station and reject unwanted adjacent stations (interference). The IF amplifiers largely determine a receiver's selectivity.
- **Fidelity:** The ability of the receiver to accurately reproduce all the frequencies present in the original modulating signal without distortion. High fidelity implies better sound quality.
- **Image Frequency Rejection:** In superheterodyne receivers, an unwanted frequency called the image frequency ($f_{image} = f_{RF} + 2f_{IF}$) can also mix with the local oscillator to produce the intermediate frequency (f_{IF}). Good image frequency rejection means the receiver can successfully block this unwanted signal before it reaches the mixer. This is typically achieved by the tuned circuits in the RF amplifier stage.

2.8 Summary

In this chapter, we studied the fundamental concepts of Amplitude Modulation (AM). We learned about the need for modulation, different modulation types, mathematical representation of AM, and its time and frequency domain analysis. We explored the power relations in AM, highlighting the efficiency of SSB transmission. Finally, we discussed the generation of DSB-SC using a balanced modulator, demodulation using an envelope detector, and the comprehensive working of a superheterodyne receiver along with its key characteristics.

CHAPTER 3

Frequency Modulation (FM) Communication

In this chapter we explore the concepts of angle modulation techniques specifically Frequency Modulation (FM) and Phase Modulation (PM). These modulation schemes offer significant advantages over Amplitude Modulation (AM) especially in terms of noise immunity and signal fidelity.

3.1 Frequency Modulation

Frequency Modulation (FM) is a form of angle modulation in which the frequency of the carrier signal is varied in proportion to the instantaneous amplitude of the modulating (message) signal. The amplitude of the carrier remains constant.

3.1.1 Mathematical Representation of FM Signal

Let the modulating signal be $m(t) = A_m \cos(2\pi f_m t)$ and the unmodulated carrier signal be $c(t) = A_c \cos(2\pi f_c t)$. In FM the instantaneous frequency f_i of the carrier changes linearly with the modulating signal:

$$f_i = f_c + k_f m(t) = f_c + k_f A_m \cos(2\pi f_m t)$$

where k_f is the frequency sensitivity of the modulator (Hz/V).

The maximum change in instantaneous frequency from the unmodulated carrier frequency is called the **Frequency Deviation** (Δf):

$$\Delta f = k_f A_m$$

The general mathematical expression for an FM wave is:

$$s(t) = A_c \cos\left(2\pi f_c t + 2\pi k_f \int m(t) dt\right)$$

For a sinusoidal modulating signal:

$$s(t) = A_c \cos(2\pi f_c t + m_f \sin(2\pi f_m t))$$

where $m_f = \frac{\Delta f}{f_m}$ is the modulation index.

Methodology:

Modulation Index and Deviation Ratio The **Modulation Index** (m_f) in FM is defined as the ratio of maximum frequency deviation to the modulating frequency:

$$m_f = \frac{\Delta f}{f_m}$$

The **Deviation Ratio** is the worst-case modulation index. It is defined as the ratio of the maximum allowable frequency deviation to the maximum modulating frequency.

$$\text{Deviation Ratio} = \frac{\Delta f_{max}}{f_{m(max)}}$$

Worked Example:

Calculation of Modulation Index An FM signal has a frequency deviation of 50 kHz and a modulating frequency of 5 kHz. Calculate the modulation index.
Solution: Given: Frequency deviation $\Delta f = 50$ kHz
Modulating frequency $f_m = 5$ kHz
The modulation index m_f is given by:

$$m_f = \frac{\Delta f}{f_m} = \frac{50 \times 10^3}{5 \times 10^3} = 10$$

3.1.2 Use of Bessel's Function and Frequency Spectrum

Unlike AM which has only two sidebands an FM wave contains theoretically an infinite number of sidebands. The amplitude of the carrier and these sidebands depend on the modulation index and are given by Bessel functions of the first kind $J_n(m_f)$.

The FM signal can be expanded as:

$$s(t) = A_c \sum_{n=-\infty}^{\infty} J_n(m_f) \cos(2\pi(f_c + nf_m)t)$$

Where:

- $J_0(m_f)$ determines the amplitude of the carrier.
- $J_1(m_f)J_2(m_f)\dots$ determine the amplitudes of the first second and subsequent sidebands.

3.1.3 Representation in Time and Frequency Domain

In the time domain an FM signal appears as a constant amplitude wave whose zero-crossings vary. When the modulating signal is at its positive peak the frequency is maximum; when it is at its negative peak the frequency is minimum.

In the frequency domain the spectrum consists of a carrier at f_c and an infinite number of sideband components at frequencies $f_c \pm nf_m$ spaced f_m apart.

3.1.4 Types of Frequency Modulation

Based on the modulation index FM is classified into two types:

1. **Narrow Band FM (NBFM):** When the modulation index $m_f \ll 1$ (usually $m_f < 0.5$). The bandwidth is approximately $2f_m$ similar to AM. It is used in mobile communication.
2. **Wide Band FM (WBFM):** When the modulation index $m_f \gg 1$. It requires a larger bandwidth and is used in FM broadcasting and TV audio. The bandwidth is calculated using Carson's rule: $BW \approx 2(\Delta f + f_m)$.

3.2 Phase Modulation

3.2.1 Definition

Phase Modulation (PM) is another form of angle modulation where the phase of the carrier signal is varied in accordance with the instantaneous amplitude of the modulating signal. The amplitude and the frequency of the unmodulated carrier remain constant.

Mathematical expression for PM:

$$s(t) = A_c \cos(2\pi f_c t + k_p m(t))$$

where k_p is the phase sensitivity (radians/V).

3.2.2 Waveforms

In PM the time-domain waveform looks very similar to FM. However the phase of the carrier changes directly with the message signal. A sudden change in the modulating signal amplitude causes a sudden phase shift in the PM wave which translates to a sharp change in instantaneous frequency.

3.2.3 Applications

- Widely used in digital modulation techniques such as Phase Shift Keying (PSK) and its variants (BPSK QPSK).
- Used in some specialized communication and radar systems.

3.3 Pre-emphasis and De-emphasis

In FM systems the noise power spectral density increases with frequency. This means that higher frequency components of the message signal are more affected by noise than lower frequency components. To counteract this effect pre-emphasis and de-emphasis are used.

Methodology:

Pre-emphasis and De-emphasis Networks **Pre-emphasis:** It is the process of artificially boosting the amplitude of the higher frequency components of the modulating signal before modulation. This is done at the transmitter using a high-pass filter circuit.
De-emphasis: It is the inverse process performed at the receiver. After demodulation the signal is passed through a low-pass filter to attenuate the high-frequency components restoring the original signal and simultaneously reducing high-frequency noise.

3.4 Generation Techniques for FM Wave

FM signals can be generated using direct or indirect methods. A modern and highly stable direct method uses a Phase Locked Loop (PLL).

3.4.1 Phase Locked Loop FM Modulator

A PLL can be configured to generate an FM signal. The basic components include a Phase Detector a Low Pass Filter and a Voltage Controlled Oscillator (VCO). When a modulating signal is applied to the control input of the VCO the frequency of the oscillator output changes in accordance with the modulating voltage. By locking this VCO to a stable reference oscillator using the PLL loop high frequency stability is achieved alongside direct frequency modulation.

3.5 Basics and Types of FM Demodulators

FM demodulation (or detection) is the process of recovering the original modulating signal from the frequency variations of the FM wave.

3.5.1 Ratio Detector

The ratio detector is a widely used FM discriminator. It converts frequency variations into amplitude variations and then detects the amplitude using diode envelope detectors.

- It consists of a transformer with a center-tapped secondary two diodes and a stabilizing capacitor.
- An advantage of the ratio detector is that it is relatively insensitive to amplitude variations (AM noise) so it does not strictly require a preceding limiter stage.

3.5.2 PLL as FM Demodulator

A Phase Locked Loop (PLL) can also be used as a high-performance FM demodulator.

- The FM signal is applied to the phase detector of the PLL.
- As the input frequency changes the phase detector produces an error voltage that passes through the loop filter.
- This error voltage forces the VCO to track the instantaneous frequency of the input FM signal.
- The error voltage itself is a replica of the original modulating signal. Hence the output of the loop filter acts as the demodulated signal.

3.6 FM Receiver

The most common architecture for an FM receiver is the superheterodyne receiver.

3.6.1 Block Diagram and Working

The standard block diagram of an FM superheterodyne receiver consists of:

1. **RF Amplifier:** Amplifies the weak received FM signal and provides initial selectivity.
2. **Mixer and Local Oscillator:** Down-converts the RF signal to an Intermediate Frequency (IF). Standard FM IF is 10.7 MHz.
3. **IF Amplifier:** Provides most of the receiver's gain and selectivity.
4. **Amplitude Limiter:** Clips off any amplitude variations (noise) ensuring a constant amplitude signal reaches the discriminator.
5. **FM Demodulator (Discriminator):** Extracts the audio signal from the FM wave.
6. **De-emphasis Network:** Restores the original amplitude-frequency response.
7. **Audio Amplifier:** Amplifies the recovered audio signal to drive a speaker.

3.7 Comparison of AM FM and PM

The differences between Amplitude Modulation (AM) Frequency Modulation (FM) and Phase Modulation (PM) are summarized below.

Table 3.1: Comparison of AM FM and PM

Parameter	**AM**	**FM**	**PM**
Modulated Parameter	Amplitude varies	Frequency varies	Phase varies
Amplitude of Wave	Varies	Constant	Constant
Noise Immunity	Poor	Very Good	Very Good
Bandwidth	Low ($2f_m$)	High (depends on m_f)	High (depends on k_p)
Transmitted Power	Varies	Constant	Constant
Complexity	Simple	Complex	Complex
Application	AM broadcasting video in TV	FM broadcasting audio in TV	Digital communication PSK

Worked Example:

Comparison of Bandwidth Determine the bandwidth required for an AM signal and a WBFM signal if the maximum modulating frequency is 15 kHz and the maximum frequency deviation for FM is 75 kHz.
Solution: For AM:

$$\text{Bandwidth}_{AM} = 2 \times f_m = 2 \times 15 \text{ kHz} = 30 \text{ kHz}$$

For WBFM (using Carson's rule):

$$\text{Bandwidth}_{FM} = 2 \times (\Delta f + f_m) = 2 \times (75 + 15) \text{ kHz} = 2 \times 90 \text{ kHz} = 180 \text{ kHz}$$

This clearly shows that FM requires much larger bandwidth than AM.

CHAPTER 4

Sampling Theory and Waveform Coding

4.1 Statement and Proof of Sampling Theorem

The sampling theorem bridges the gap between continuous-time analog signals and discrete-time signals. It is a fundamental principle in digital communication systems, ensuring that no information is lost during the conversion from analog to digital form.

4.1.1 Statement

The sampling theorem states that a continuous-time signal $x(t)$ which is band-limited to f_m Hertz can be completely represented by its samples and recovered without error if the sampling frequency f_s is at least twice the maximum frequency component f_m present in the signal. Mathematically, this condition is expressed as:

$$f_s \geq 2f_m$$

4.1.2 Proof of Sampling Theorem

Let $x(t)$ be a continuous-time signal band-limited to f_m. The frequency spectrum of $x(t)$ is $X(f)$ and it is zero for $|f| > f_m$. Let this signal be multiplied by an ideal impulse train $\delta_T(t)$, which consists of unit impulses spaced by the sampling interval T_s.

The impulse train is represented as:

$$\delta_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s)$$

The sampled signal $x_s(t)$ is the product of $x(t)$ and $\delta_T(t)$:

$$x_s(t) = x(t) \cdot \delta_T(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t - nT_s)$$

Taking the Fourier Transform of the sampled signal $x_s(t)$, we apply the property that multiplication in the time domain corresponds to convolution in the frequency domain. The Fourier transform of the impulse train is another impulse train in the frequency domain:

$$F\{\delta_T(t)\} = f_s \sum_{n=-\infty}^{\infty} \delta(f - nf_s)$$

where $f_s = \frac{1}{T_s}$.

The frequency spectrum of the sampled signal $X_s(f)$ is obtained by convolving $X(f)$ with the Fourier transform of the impulse train:

$$X_s(f) = X(f) * \left(f_s \sum_{n=-\infty}^{\infty} \delta(f - nf_s)\right)$$

$$X_s(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)$$

This equation shows that the spectrum of the sampled signal consists of the original spectrum $X(f)$ scaled by f_s and repeated infinitely at frequency intervals of $\pm f_s, \pm 2f_s, \ldots$. To reconstruct the original signal without any distortion (aliasing), these shifted spectra must not overlap. The condition for no overlap between adjacent spectra is:

$$f_s - f_m \geq f_m \implies f_s \geq 2f_m$$

If this condition is met, the original signal can be perfectly recovered by passing the sampled signal through an ideal low-pass filter with a cutoff frequency f_c such that $f_m \leq f_c \leq (f_s - f_m)$.

4.2 Nyquist Rate and Interval

4.2.1 Nyquist Rate

The Nyquist rate is the theoretical minimum sampling rate required to avoid aliasing and enable perfect reconstruction of the analog signal. It is exactly twice the maximum frequency component present in the analog signal.

$$f_{Nyquist} = 2f_m$$

If a signal is sampled at exactly the Nyquist rate ($f_s = 2f_m$), it is referred to as critical sampling.

4.2.2 Nyquist Interval

The Nyquist interval is the maximum allowed time interval between successive samples to guarantee perfect reconstruction of the signal. It is simply the reciprocal of the Nyquist rate.

$$T_{Nyquist} = \frac{1}{2f_m}$$

Methodology:

Calculating Nyquist Rate and Interval To find the Nyquist rate and interval for a given band-limited signal:

1. Identify the maximum frequency component f_m present in the signal spectrum.
2. Multiply f_m by 2 to find the Nyquist rate $f_{Nyquist} = 2f_m$.
3. Take the reciprocal of the Nyquist rate to determine the Nyquist interval $T_{Nyquist} = 1/f_{Nyquist}$.

Worked Example:

Nyquist Rate Calculation Calculate the Nyquist rate and Nyquist interval for the signal $x(t) = 10\cos(2000\pi t) + 5\sin(4000\pi t)$.

Solution: First, we find the frequencies of the individual components of the signal. The angular frequency of the first term ω_1 is 2000π, so its frequency in Hertz is $f_1 = \frac{2000\pi}{2\pi} = 1000$ Hz. The angular frequency of the second term ω_2 is 4000π, so its frequency is $f_2 = \frac{4000\pi}{2\pi} = 2000$ Hz.

The maximum frequency component f_m is the largest of these individual frequencies:

$$f_m = \max(1000 \text{ Hz}, 2000 \text{ Hz}) = 2000 \text{ Hz}$$

The Nyquist rate is calculated as:

$$f_{Nyquist} = 2f_m = 2 \times 2000 = 4000 \text{ Hz (or 4 kHz)}$$

The Nyquist interval is:

$$T_{Nyquist} = \frac{1}{f_{Nyquist}} = \frac{1}{4000} = 0.25 \text{ ms}$$

4.3 Aliasing Error and Sampling Types

4.3.1 Aliasing Error

Aliasing is an unwanted phenomenon that occurs when a continuous-time signal is sampled at a rate lower than the Nyquist rate ($f_s < 2f_m$). In the frequency domain, the shifted spectra $X(f - nf_s)$ overlap with each other. Due to this overlap, the high-frequency components of the original signal masquerade as low-frequency components in the sampled signal. This overlap causes irreversible distortion, and the original analog signal cannot be perfectly reconstructed from its samples.

To prevent aliasing in practical systems, two precautions are generally taken:

- Ensure the sampling frequency strictly satisfies $f_s \geq 2f_m$.
- Pass the analog signal through an anti-aliasing filter (a low-pass filter) before sampling to strictly limit its bandwidth to f_m and remove high-frequency noise.

4.3.2 Types of Sampling Rates

Based on how the sampling frequency f_s compares to the maximum signal frequency f_m, sampling is classified into three categories:

1. Under Sampling ($f_s < 2f_m$) When the sampling rate is less than the Nyquist rate, the adjacent spectral bands overlap. This causes aliasing error, and the original signal is irrevocably distorted. Under sampling is generally avoided in standard waveform coding for baseband signals.

2. Critical Sampling ($f_s = 2f_m$) When the sampling rate is exactly equal to the Nyquist rate, the adjacent spectra in the frequency domain just touch each other without overlapping. While theoretically the signal can be recovered, it requires an ideal low-pass filter (brick-wall filter) with an infinitely sharp cutoff to isolate the original spectrum. Since ideal filters cannot be physically realized, critical sampling is rarely used in practice.

3. Over Sampling ($f_s > 2f_m$) When the sampling rate is strictly greater than twice the maximum frequency, the spectra $X(f - nf_s)$ are well-separated with a gap known as a guard band between them. This is the most practical sampling method as the guard band accommodates the gradual roll-off of practical non-ideal low-pass filters used for signal reconstruction.

4.4 Ideal Natural and Flat Top Sampling

Sampling techniques can also be classified based on the shape and characteristics of the sampling pulses used.

4.4.1 Ideal Sampling

Also known as instantaneous or impulse sampling, this theoretical method uses a train of ideal impulse functions $\delta(t)$ to sample the analog signal. The sampled signal is a sequence of weighted impulses whose strengths correspond to the instantaneous amplitude of the continuous signal at the exact sampling instants. Since true mathematical impulses have zero time duration and infinite amplitude, ideal sampling cannot be practically implemented. However, it serves as the essential theoretical foundation for analyzing sampling.

4.4.2 Natural Sampling

In natural sampling, the continuous signal is sampled using a periodic train of rectangular pulses with finite width τ and a constant amplitude A. During the pulse duration τ, the top of the sampling pulse follows the exact contour (shape) of the original analog signal. The sampled signal $x_{ns}(t)$ is obtained by directly multiplying the original signal $x(t)$ with the rectangular pulse train using a switching circuit (such as a FET switch). Natural sampling is practically realizable but relatively complex to process digitally because the pulse amplitude is not constant over its duration.

4.4.3 Flat Top Sampling

In flat top sampling, the sampled signal consists of rectangular pulses of finite width τ, but unlike natural sampling, the amplitude (the top) of each pulse is held rigidly constant throughout the pulse duration. The constant amplitude is equal to the instantaneous value of the analog signal precisely at the beginning of the sampling instant. This technique is implemented using a Sample and Hold (S/H) circuit. Flat top sampling introduces a slight frequency distortion known as the aperture effect, which acts like a low-pass filter and mildly attenuates higher frequencies. This effect can be compensated at the receiver using an equalizer. Flat top sampling is the most widely used technique in digital communication and Pulse Code Modulation (PCM) because constant amplitude pulses are vastly easier to convert into digital codes.

4.5 Pulse Modulation Techniques

Analog pulse modulation techniques use a discrete-time continuous-amplitude signal to carry information. A periodic train of rectangular pulses acts as the carrier, and one of its parameters (amplitude, width, or position) is varied in direct proportion to the modulating signal.

4.5.1 Pulse Amplitude Modulation (PAM)

In PAM, the amplitude of each pulse in the carrier pulse train is varied proportionally to the instantaneous value of the modulating analog signal.

- The width and position of all pulses remain constant.
- It is the simplest form of pulse modulation and is conceptually similar to flat top or natural sampling.
- **Advantage:** Very simple to generate and demodulate.
- **Disadvantage:** It is highly susceptible to noise interference because noise readily alters the amplitude of the transmitted pulses.

4.5.2 Pulse Width Modulation (PWM)

In PWM (also known as Pulse Duration Modulation - PDM or Pulse Length Modulation - PLM), the width (duration) of the pulses is varied in proportion to the amplitude of the modulating signal.

- The amplitude and the position of the pulse centers usually remain constant.
- A higher positive amplitude of the modulating signal results in a correspondingly wider pulse.
- **Advantage:** Provides better noise immunity than PAM because the information is contained entirely in the pulse width, allowing amplitude limiters at the receiver to clip off amplitude noise.
- **Disadvantage:** Requires a wider variable transmission bandwidth, and the transmission power varies since wider pulses consume significantly more power.

4.5.3 Pulse Position Modulation (PPM)

In PPM, the relative position of each pulse with respect to its unmodulated time slot is shifted in accordance with the amplitude of the modulating signal.

- The amplitude and width of all transmitted pulses remain strictly constant.
- PPM is often generated derived from PWM. The trailing edge of a PWM pulse is used to trigger a very short-duration PPM pulse.
- **Advantage:** Offers excellent noise immunity. Since both amplitude and width are constant, the transmitter requires constant, low average power, making it highly power-efficient.
- **Disadvantage:** Synchronization between the transmitter and receiver is strictly required to determine the precise unmodulated pulse positions for accurate demodulation.

4.6 Concept of Quantization

Quantization is the pivotal process of converting a continuous-amplitude signal into a discrete-amplitude signal. While sampling discretizes the signal in the time domain, quantization discretizes the signal in the amplitude domain.

In the quantization process, the continuous range of possible amplitude values is partitioned into a finite number of discrete, predetermined levels. Each sampled amplitude obtained from the continuous signal is approximated (or rounded off) to the nearest available discrete level. The unavoidable difference between the actual unquantized sample value and the assigned quantized value is called **quantization error** or **quantization noise**. This noise is inherent to the quantization process and cannot be completely eliminated, though it can be minimized.

For instance, if an analog signal ranges from 0 to 5V and we employ a 3-bit quantizer, there will be $2^3 = 8$ discrete quantization levels available. A sampled analog value of 3.2V might be approximated to a quantization level of 3.1V, resulting in a quantization error of 0.1V.

The separation between adjacent quantization levels is called the step size (Δ) and for uniform quantization, it is given by:

$$\Delta = \frac{V_{max} - V_{min}}{L}$$

where $V_{max} - V_{min}$ represents the peak-to-peak dynamic range of the signal, and $L = 2^n$ is the total number of quantization levels for an n-bit binary quantizer.

4.7 Classification of Quantization

Quantization strategies are broadly classified into two main types based on how the step sizes are distributed across the dynamic range.

4.7.1 Uniform Quantization

In uniform quantization, the step size Δ remains strictly constant throughout the entire dynamic range of the signal. The quantization levels are equally spaced from one another.

- **Advantage:** Straightforward and simple to implement in hardware.
- **Disadvantage:** The Signal-to-Quantization-Noise Ratio (SQNR) is poor for weak (low amplitude) signals and unnecessarily high for strong signals. This makes uniform quantization highly inefficient for signals characterized by a large dynamic range, such as human speech, where low-amplitude sounds occur much more frequently than loud sounds.

Uniform quantizers are generally modeled in two ways:

- **Mid-tread Quantizer:** Possesses a distinct quantization level situated exactly at zero volts.
- **Mid-riser Quantizer:** Features a step transition (a riser) situated exactly at zero volts.

4.7.2 Non-uniform Quantization

In non-uniform quantization, the step size varies deliberately depending on the amplitude of the signal. The step sizes are kept very small for weak (low amplitude) signals to minimize quantization noise, and large for strong (high amplitude) signals where larger errors are less perceptible.

- **Advantage:** Provides a nearly constant and optimal SQNR across the entire dynamic range of the signal, which drastically improves the fidelity of audio and voice communications.
- **Implementation:** Non-uniform quantization is almost exclusively implemented using a technique called **Companding** (a portmanteau of Compressing and Expanding). The analog signal is non-linearly compressed at the transmitter (which amplifies weak signals much more than strong ones) before undergoing standard uniform quantization. At the receiver, an inverse expansion process restores the original dynamic range. The two globally standardized companding curves are the ţ-law (predominantly used in North America and Japan) and the A-law (used in Europe and the rest of the world).

4.8 PCM Transmitter and Receiver

Pulse Code Modulation (PCM) is the predominant standard method utilized to digitally represent and transmit sampled analog signals. It forms the critical foundation for modern digital voice telephony, digital audio, and many other digital communication systems.

4.8.1 PCM Transmitter

The PCM transmitter is responsible for converting a continuous analog signal into a robust digital bit stream. It comprises four essential functional blocks:

1. **Low Pass Filter (Anti-aliasing Filter):** The raw continuous analog signal first traverses an anti-aliasing low-pass filter to rigidly restrict the signal bandwidth to f_m. This crucial step prevents high-frequency components and out-of-band noise from causing aliasing distortion during the subsequent sampling stage.

2. **Sampler (Sample and Hold Circuit):** The band-limited analog signal is sampled at a frequency $f_s \geq 2f_m$. A Sample and Hold (S/H) circuit performs flat-top sampling, securely holding the amplitude steady while the subsequent analog-to-digital conversion takes place. This generates a discrete-time PAM signal.

3. **Quantizer:** The discrete-time continuous-amplitude PAM samples are then sequentially fed into the quantizer. It approximates each continuous sample to the nearest discrete quantization level, transforming the continuous amplitude into a discrete amplitude.

4. **Encoder:** The encoder systematically assigns a unique binary code (a structured sequence of bits) to each individual quantized level. For an L-level quantizer, the encoder outputs an n-bit binary word per sample, where $L = 2^n$. The final output of the transmitter is a serialized digital binary pulse stream.

4.8.2 PCM Receiver

The PCM receiver performs the inverse operations to flawlessly reconstruct the original continuous analog signal from the corrupted incoming digital bit stream.

1. **Regenerative Repeater:** Over long transmission distances or noisy channels, the digital signal severely attenuates and accumulates noise. Regenerative repeaters strategically placed along the transmission path detect, reshape, and fully reconstruct the binary pulses, effectively stripping away accumulated noise before it can cause irrecoverable bit errors.

2. **Decoder:** The receiver captures the incoming serial bit stream and carefully groups the bits back into original n-bit words. The decoder mathematically converts each binary word back into its corresponding discrete-amplitude voltage level. This accurately regenerates a quantized PAM signal.

3. **Reconstruction Filter (Low Pass Filter):** The regenerated quantized PAM signal is finally passed through a low-pass reconstruction filter featuring a precise cutoff frequency equal to f_m. This final filter smooths the harsh staircase-like PAM signal, mathematically interpolating between the discrete samples to reconstruct a continuous-time analog signal that closely mirrors the original input.

Methodology:

Advantages of PCM Systems Pulse Code Modulation provides several revolutionary advantages over analog systems:

- Exceptionally high immunity to channel noise and interference.
- The digital signal can be perfectly regenerated over immense distances using repeaters.
- The uniform digital format easily allows for Time Division Multiplexing (TDM) of diverse signals like voice video and data.
- Digital data can be readily encrypted for highly secure communication and error-correction coded for ultimate reliability.

CHAPTER 5

Electromagnetic Wave and antenna

5.1 Introduction

Communication systems rely heavily on the transmission of signals over long distances. For wireless communication, signals are transmitted and received through open space using electromagnetic (EM) waves. An antenna serves as the crucial interface between the electronic circuits of the transmitter or receiver and the free space medium. In this chapter, we will study the properties of electromagnetic waves, fundamental concepts of antennas, their key parameters, and different types of antennas including their working principles and applications.

5.2 Properties of EM (Electromagnetic) Wave

An electromagnetic (EM) wave is created as a result of vibrations between an electric field and a magnetic field. In a vacuum, these waves travel at the speed of light.

Methodology:

Properties of Electromagnetic Waves EM waves possess several important properties that dictate how they behave in various media and in free space:

- **Transverse Nature:** The electric field (**E**) and magnetic field (**H**) are perpendicular to each other and both are perpendicular to the direction of wave propagation.
- **Speed of Propagation:** In a vacuum, all EM waves travel at the speed of light, denoted by $c \approx 3 \times 10^8$ m/s.
- **No Medium Required:** Unlike sound waves, EM waves do not require a material medium to propagate; they can travel through a vacuum.
- **Wavelength and Frequency Relation:** The speed, frequency (f), and wavelength (λ) of an EM wave are related by the equation $c = f\lambda$.
- **Energy Transfer:** EM waves carry energy and momentum which can be transferred to objects they interact with.
- **Reflection and Refraction:** EM waves undergo reflection, refraction, interference, diffraction, and polarization similar to light waves.

Worked Example:

Calculating Wavelength An FM radio station broadcasts at a frequency of 98.3 MHz. Calculate the wavelength of the electromagnetic wave produced by the station. Assume the speed of light $c = 3 \times 10^8$ m/s.
Solution: Given: Frequency $f = 98.3$ MHz $= 98.3 \times 10^6$ Hz Speed of light $c = 3 \times 10^8$ m/s
The relationship between wavelength, frequency, and speed is:

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

Substitute the given values:

$$\lambda = \frac{3 \times 10^8}{98.3 \times 10^6}$$

$$\lambda \approx \frac{300}{98.3} \approx 3.05 \text{ meters}$$

The wavelength of the radio wave is approximately 3.05 meters.

5.3 Antenna Fundamentals

An antenna is a metallic structure (often a wire or collection of wires) that acts as a transducer, converting electrical power into electromagnetic waves and vice versa. It is used in systems such as radio broadcasting, broadcast television, two-way radio, and communication receivers.

5.3.1 Resonant and Nonresonant Antennas

Antennas can be broadly classified into resonant and nonresonant types based on their electrical length and the standing wave patterns they support.

Resonant Antennas

- A resonant antenna is an antenna whose physical length is specifically related to the wavelength of the transmitted or received signal (typically multiples of a half-wavelength, $\lambda/2$).
- They are open-ended, meaning the current drops to zero at the ends, producing standing waves of voltage and current along the antenna.
- They are highly efficient but operate over a narrow band of frequencies.
- Example: Half-wave dipole antenna.

Nonresonant Antennas

- A nonresonant antenna (or traveling-wave antenna) is terminated with a resistive load equal to its characteristic impedance, preventing signal reflection from the end.
- Only forward-traveling waves exist; there are no standing waves.
- They operate over a much broader range of frequencies (wide bandwidth) but are generally less efficient than resonant antennas because power is dissipated in the terminating resistor.
- Example: Rhombic antenna.

5.3.2 Ideal Antenna

An ideal antenna is a theoretical model known as an isotropic radiator.

- It is defined as a hypothetical point source that radiates electromagnetic energy equally in all directions spherically.
- The radiation pattern is a perfect sphere.
- An isotropic antenna does not exist in reality, but it is used as a standard reference to measure the performance (such as directivity and gain) of practical antennas.

5.3.3 Principle of Transmitting and Receiving Antenna

The principle of operation of an antenna relies on the laws of electromagnetism (Maxwell's equations).

Transmitting Antenna Principle: When a high-frequency alternating current from a transmitter is applied to the terminals of a transmitting antenna, the moving electrical charges (electrons) in the antenna element create alternating electric and magnetic fields. According to Maxwell's equations, a changing electric field creates a changing magnetic field, and vice versa. This mutual interaction results in the generation of an electromagnetic wave that propagates away from the antenna into space.

Receiving Antenna Principle: When an electromagnetic wave travels through space and passes across a receiving antenna, the alternating electric and magnetic fields of the wave exert a force on the electrons in the antenna material. This causes the electrons to move back and forth, inducing an alternating electrical current in the antenna that matches the frequency of the incoming wave. This weak current is then fed to the receiver circuit for amplification and demodulation.

Antenna Reciprocity: The theorem of reciprocity states that the characteristics of an antenna, such as its radiation pattern, directivity, and gain, are identical whether the antenna is being used for transmitting or receiving.

5.4 Antenna Parameters

To evaluate and compare the performance of different antennas, specific parameters are defined.

5.4.1 Radiation Pattern

The radiation pattern (or antenna pattern) is a graphical representation of the radiation properties of the antenna as a function of space coordinates.

- In most cases, it is determined in the far-field region and is represented as a function of directional coordinates (azimuth and elevation angles).
- It shows how the antenna radiates power in various directions.
- Important parts of a radiation pattern include the major lobe (main beam with maximum radiation) and minor lobes (side lobes and back lobes).

5.4.2 Polarization

Polarization of an antenna refers to the orientation of the electric field (**E**-field) of the electromagnetic wave radiated by the antenna.

- **Linear Polarization:** The electric field vector oscillates in a single plane. It can be horizontal (e.g., standard TV broadcasting) or vertical (e.g., mobile communications).
- **Circular Polarization:** The electric field vector rotates in a circle as the wave propagates.
- **Elliptical Polarization:** The electric field vector traces an ellipse.
- For optimum reception, the receiving antenna must have the same polarization as the transmitting antenna.

5.4.3 Bandwidth and Beam Width

Bandwidth: The bandwidth of an antenna is the range of frequencies over which the antenna can operate efficiently and correctly meet its specified parameters (like impedance, gain, or radiation pattern). It is often specified as a percentage of the center frequency.

Beam Width: Also known as Half-Power Beamwidth (HPBW), it is the angular separation between two points on the major lobe of the radiation pattern where the radiated power is half (3 dB less than) the maximum value. It indicates the directivity of the antenna; a smaller beam width means a highly directive antenna.

5.4.4 Antenna Resistance

The input impedance of an antenna at its terminals has a resistive component called the antenna resistance. It consists of two parts:

- **Radiation Resistance (R_r):** A fictitious resistance that accounts for the power radiated by the antenna into space. If P is the radiated power and I is the RMS current, $P = I^2 R_r$.
- **Loss Resistance (R_l):** The actual ohmic resistance of the antenna conductors and dielectric losses, which dissipate power as heat.

Total antenna resistance $R_{ant} = R_r + R_l$.

5.4.5 Directivity, Power Gain and Antenna Gain

Directivity (D): Directivity is the ratio of the radiation intensity in a given direction from the antenna to the radiation intensity averaged over all directions. The average radiation intensity is equal to the total power radiated by the antenna divided by 4π. It measures how concentrated the antenna's radiation is in a particular direction compared to an isotropic radiator.

Power Gain (G): Power gain is closely related to directivity but also accounts for the efficiency of the antenna. It is the ratio of the radiation intensity in a given direction to the radiation intensity that would be obtained if the power accepted by the antenna were radiated isotropically.

$$G = \eta \cdot D$$

where η is the antenna efficiency ($\eta = \frac{R_r}{R_r + R_l}$).

Worked Example:

Antenna Gain Calculation An antenna has a radiation resistance of 72 Ω and a loss resistance of 8 Ω. If its directivity is 15, calculate its radiation efficiency and maximum power gain.
Solution: Given: Radiation resistance $R_r = 72\ \Omega$ Loss resistance $R_l = 8\ \Omega$ Directivity $D = 15$
1. Radiation Efficiency (η):

$$\eta = \frac{R_r}{R_r + R_l} = \frac{72}{72 + 8} = \frac{72}{80} = 0.9$$

Percentage efficiency = 90%
2. Maximum Power Gain (G):

$$G = \eta \times D = 0.9 \times 15 = 13.5$$

To express gain in decibels (dB):

$$G_{dB} = 10 \log_{10}(13.5) \approx 11.3 \text{ dB}$$

The antenna has an efficiency of 90% and a power gain of 13.5 (or 11.3 dB).

5.5 Dipole Antenna and Radiation Pattern

5.5.1 Dipole Antenna

The dipole antenna is one of the simplest and most widely used types of antennas. It consists of two straight metallic conductors (poles) oriented end-to-end, with the feed line connected to the gap between them.

- The most common form is the **half-wave dipole**, where the total length of the antenna is equal to half the wavelength ($\lambda/2$) of the operating frequency.
- It is a resonant antenna.
- The current distribution along a half-wave dipole is maximum at the center (feed point) and zero at the ends. Voltage is zero at the center and maximum at the ends.
- The radiation resistance of a standard half-wave dipole in free space is approximately 73 Ω.

5.5.2 Radiation Pattern for Unidirectional, Bidirectional and Omnidirectional Antenna

Methodology:

Types of Radiation Patterns The spatial distribution of radiated energy defines the directivity of the antenna:

1. **Omnidirectional Radiation Pattern:**
   - Radiates power equally in all directions in one plane (usually the horizontal plane), while the radiated power varies with the elevation angle.
   - The pattern looks like a doughnut.
   - A vertical half-wave dipole produces an omnidirectional pattern in the horizontal plane. Ideal for broadcasting where receivers are scattered in all directions.
2. **Bidirectional Radiation Pattern:**
   - Radiates power maximally in two opposite directions.
   - The pattern has two main lobes that are 180 degrees apart.
   - A horizontal half-wave dipole oriented along the X-axis produces maximum radiation along the positive and negative Y-axes, making it bidirectional in the XY plane.
3. **Unidirectional Radiation Pattern:**
   - Radiates power predominantly in one specific direction.
   - Achieved by using reflector elements or phased arrays to suppress radiation in unwanted directions and concentrate it in a single forward lobe.
   - Highly directive, making it suitable for point-to-point communication links.
   - Examples include Yagi-Uda antenna and parabolic dish antennas.

5.6 Specific Antennas

Let us explore the working principles, construction, radiation patterns, and applications of some common antennas.

5.6.1 Loop Antenna

Construction: A loop antenna consists of a circular, square, or rectangular coil of wire. It can be a single turn or multiple turns of wire. The dimensions of the loop are usually much smaller than a wavelength (small loop) for typical low-frequency applications, though large loops (perimeter $\approx \lambda$) are also used.

Working Principle: The loop operates on the principle of electromagnetic induction. When an alternating magnetic field from an incident EM wave passes through the loop, it induces a voltage in the wire proportional to the rate of change of magnetic flux.

Radiation Pattern: For a small loop antenna, the radiation pattern is a figure-eight shape (bidirectional), very similar to that of a half-wave dipole, but the maximum radiation occurs in the plane of the loop.

Applications:

- Direction finding equipment (due to the sharp nulls perpendicular to the loop plane).
- AM broadcast receivers (built-in ferrite rod loop antennas).
- RFID tags and near-field communication (NFC) systems.

5.6.2 Folded Dipole Antenna

Construction: A folded dipole is formed by joining two half-wave dipoles at their ends to form a closed continuous loop. The feed point is at the center of one of the conductors. The total perimeter is essentially one full wavelength, making its length $\lambda/2$.

Working Principle: It works similarly to a standard dipole but behaves as two parallel conductors. The folded structure acts as an impedance transformer. The input impedance of a standard half-wave dipole is about 73 Ω, whereas a folded dipole made of equal diameter wires has an input impedance of $4 \times 73\ \Omega \approx 292\ \Omega$ (often taken as 300 Ω).

Radiation Pattern: The radiation pattern is identical to that of a standard half-wave dipole: a bidirectional figure-eight pattern perpendicular to the antenna axis.

Applications:

- Used extensively as the driven element in Yagi-Uda antennas because its higher impedance matches well with standard 300 Ω twin-lead transmission lines.
- FM and TV broadcast receiving antennas.

5.6.3 Yagi-Uda Antenna

Construction: Invented by Shintaro Uda and Hidetsugu Yagi, it is a highly directional array antenna. It consists of multiple parallel linear elements arranged in a line:

- **Driven Element:** Usually a folded dipole (length $\approx 0.47\lambda$) connected to the transmitter or receiver.
- **Reflector:** Placed behind the driven element. It is slightly longer than the driven element (length $\approx 0.5\lambda$). There is typically only one reflector.
- **Directors:** Placed in front of the driven element. They are slightly shorter than the driven element (length $\approx 0.43\lambda$ to 0.45λ). There can be multiple directors to increase gain.

Working Principle: The driven element radiates EM waves. These waves induce currents in the parasitic elements (reflector and directors). The reflector re-radiates the signal with a phase shift that cancels radiation behind the antenna and reinforces it in the forward direction. The directors further focus the wave in the forward direction, functioning as directors of the electromagnetic energy.

Radiation Pattern: It has a highly unidirectional radiation pattern with a strong major lobe in the forward direction (towards the directors) and very small minor lobes.

Applications:

- VHF and UHF television reception (classic rooftop TV antennas).
- Point-to-point communication links.
- Amateur radio (HAM) communication.

5.6.4 Smart Antenna

Construction: A smart antenna (or adaptive array antenna) is not just a single mechanical antenna, but an array of antenna elements combined with sophisticated digital signal processing (DSP) algorithms.

Working Principle: Smart antennas automatically adjust their radiation pattern in response to the signal environment. The DSP system analyzes the incoming signals, identifies the direction of the desired user and the directions of interfering signals. It then adjusts the phase and amplitude of the signals sent to or received from each antenna element. This allows the antenna array to form a directed beam (main lobe) directly toward the desired user and place nulls (zero radiation) in the directions of interferers.

Types of Smart Antennas:

- **Switched Beam:** Contains predefined, fixed directional beam patterns and the system switches to the beam that gives the best reception.
- **Adaptive Array:** Continuously steers the beam and nulls dynamically in real-time tracking the users and interferers.

Radiation Pattern: Highly directive and dynamically changing unidirectional pattern, steered electronically without physically moving the antenna.

Applications:

- Cellular networks (4G LTE and 5G) to increase capacity and data rates (Massive MIMO).
- Wi-Fi systems (IEEE 802.11n/ac/ax) using beamforming technology.
- Radar and military communications.